DIE A-POSTERIORI-SCHICHTUNG

Hochrechnung, Wirksamkeit und Anwendung
bei gleicher und ungleicher Auswahlwahrscheinlichkeit

Inaugural-Dissertation
zur Erlangung des Doktorgrades einer
Hohen Rechts- und Wirtschaftswissenschaftlichen Fakultät
(Wirtschaftswissenschaftliche Abteilung)
der Eberhard-Karls-Universität zu Tübingen

vorgelegt von
Siegfried Hauser
aus Fischingen

~~1970~~

1969

DIE A-POSTERIORI-SCHICHTUNG

Hochrechnung, Wirksamkeit und Anwendung
bei gleicher und ungleicher Auswahlwahrscheinlichkeit

Inaugural-Dissertation
zur Erlangung des Doktorgrades einer
Hohen Rechts- und Wirtschaftswissenschaftlichen Fakultät
(Wirtschaftswissenschaftliche Abteilung)
der Eberhard-Karls-Universität zu Tübingen

vorgelegt von
Siegfried Hauser
aus Fischingen

1970

UNIV.
LEIDEN
BIBL.

Dekan: Prof. Dr.rer.nat. H. Strecker

Erstberichterstatter: Prof. Dr.rer.nat. H. Strecker

Zweitberichterstatter: Privatdozent Dr.rer.pol. D. Lüdeke

Tag der mündlichen Prüfung: 10.Juli 1969

V o r w o r t

Der Statistik und insbesondere den Stichprobenmethoden ist die Aufgabe gestellt, für die vielfältigsten Bereiche des täglichen Lebens und der Wissenschaft Zahlen zu beschaffen. Mit Hilfe dieser Zahlen sollen z.B. in der Wirtschaft die strukturellen Zusammenhänge durchleuchtet und Indikatoren geschaffen werden, die Änderungen im Wirtschaftsablauf aufzeigen. Um dem Zahlenbedarf auf aktuelle Weise und mit angebbarer Fehlergrenze nachkommen zu können, müssen neue Stichprobenverfahren entwickelt und bestehende verfeinert werden. Hierzu soll die vorliegende Arbeit einen Beitrag liefern.

Die Arbeit wurde angeregt durch Stichprobenuntersuchungen auf dem Gebiete der Agrarstatistik, die am Institut für Wirtschafts- und Sozialstatistik im Rahmen der statistischen Arbeit der Europäischen Wirtschaftsgemeinschaft aus methodischen Gründen durchgeführt worden sind.

Da das Inhaltsverzeichnis sehr umfangreich und deshalb etwas unübersichtlich ist, wird zunächst in einer "Inhaltsübersicht" ein grober Überblick über die wichtigsten Abschnitte gegeben. Die Feingliederung folgt dann im "Inhaltsverzeichnis".

Für das freundliche Interesse am Gelingen dieser Arbeit danke ich Herrn Professor Dr. H.Strecker herzlich.

Inhaltsübersicht

0 Notationen VIII

1 Einleitung 1

2 Voraussetzungen 4

2.1 Schätzprobleme im Allgemeinen 4

2.2 Stichproben 20

2.3 Die bedeutendsten Hochrechnungsarten neben der A-posteriori-Schichtung 30

3 A-posteriori-Schichtung 52

3.1 Prinzip dieser Methode 52

3.2 Die A-posteriori-Schichtung, ein Post-hoc-Verfahren und seine Vorteile 55

3.3 A-posteriori-Schichtung und verwandte Verfahren 59

3.4 Theorie 67

3.5 Wirksamkeit 90

3.6 Schichtenbildung 106

3.7 A-posteriori-Schichtung bei Auswahl mit ungleichen Wahrscheinlichkeiten 124

3.8 Kombination der A-posteriori-Schichtung mit anderen Hochrechnungsverfahren 131

3.9 Anwendung 137

3.10 Anwendungsbeispiele 140

4 Abschließende Bemerkungen 156

5 Literaturverzeichnis 159

Inhaltsverzeichnis

0 NOTATIONEN VIII

1 EINLEITUNG 1

2 VORAUSSETZUNGEN 4

2.1 Schätzprobleme im Allgemeinen 4
2.1.1 Geschichtlicher Rahmen 4
2.1.2 Grundlagen der Schätztheorie 7
2.1.2.1 Das Schätzen von Parametern 7
2.1.2.2 Statistische Schlußweise 8
2.1.2.3 Eigenschaften guter Schätzfunktionen 11
2.1.2.3.1 Verzerrte und unverzerrte Schätzung 11
2.1.2.3.2 Konsistenz 15
2.1.2.3.3 Effizienz (Wirksamkeit) 17
2.1.2.3.4 Suffizienz (erschöpfende Schätzungen) 19

2.2 Stichproben 20
2.2.1 Der Stichprobenplan 20
2.2.2 Die Stichprobenbasis 22
2.2.3 Das Stichprobenmodell 24
2.2.4 Die Stichprobendurchführung 26

2.3 Die bedeutenden Hochrechnungsarten neben der A-posteriori-Schichtung 30
2.3.1 Allgemeine Einteilung 30
2.3.2 Freie oder einfache Hochrechnung 31
2.3.3 Verhältnisschätzung 32
2.3.3.1 Methode 32
2.3.3.2 Formeln für Varianz und Bias 35
2.3.3.3 Bias und Standardabweichung 37
2.3.3.4 Eine Variante der Verhältnisschätzung 39
2.3.3.5 Stichprobenauswahl mit der Wahrscheinlichkeit proportional zur Hilfsvariablen x 41
2.3.3.6 Separate und kombinierte Verhältnisschätzung 42
2.3.4 Regressionsschätzung 43
2.3.4.1 Methode 43
2.3.4.2 Varianz der Regressionsschätzung 45
2.3.4.3 Bias 48
2.3.4.4 Die Methode von Abraham Wald 49
2.3.4.5 Separate und kombinierte Regressionsschätzung 50
2.3.5 Differenzenschätzung 50

3 A-POSTERIORI-SCHICHTUNG 52

3.1 Prinzip dieser Methode 52

3.2 Die A-posteriori-Schichtung, ein Post-hoc-Verfahren und seine Vorteile 55

3.3 A-posteriori-Schichtung und verwandte Verfahren 59

3.3.1 Zwei-Phasen-Stichprobe und Post-Klumpenstichprobe 59

3.3.2 A-posteriori-Schichtung und Quotenverfahren 60

3.3.3 A-posteriori-Schichtung und Gruppierungsverfahren (grouping) 63

3.4 Theorie 67

3.4.1 Erwartungstreue des Schätzwerts (bei besetzten Schichten) 67

3.4.2 Erwartungswert bei nicht besetzten Schichten 70

3.4.3 Varianz 73

3.4.4 Schätzwert der Varianz 79

3.4.5 Bias und Standardabweichung 81

3.4.6 Die A-posteriori-Schichtung bei gleich großen Schichten 88

3.5 Wirksamkeit 90

3.5.1 Theoretische Betrachtung 90

3.5.1.1 Vergleich mit der einfachen Hochrechnung und dem geschichteten Auswahlverfahren 91

3.5.1.2 Vergleich mit der Verhältnisschätzung 92

3.5.1.3 Vergleich mit der Differenzenschätzung 93

3.5.1.4 Vergleich mit der Regressionsschätzung 94

3.5.1.5 Zusammenfassende Darstellung 95

3.5.2 Rechenbeispiel zur Demonstration der Wirksamkeit 99

3.5.2.1 Einfache Hochrechnung 101

3.5.2.2 A-posteriori-Schichtung 101

3.5.2.3 Verhältnisschätzung 102

3.5.2.4 Differenzenschätzung 103

3.5.2.5 Regressionsschätzung 103

3.5.2.6 Vergleichende Zusammenstellung 104

3.6 Schichtenbildung 106

3.6.1 Allgemeine Theorie der Schichtung 106

3.6.2 Schätztheorie der Schichtung 112

3.6.2.1 Einfache Schätzung aus der Stichprobe 113

3.6.2.2 Schätzung der Anteilssätze unter Annahme einer A-priori-Verteilung 115

3.7 A-posteriori-Schichtung bei Auswahl mit ungleichen Wahrscheinlichkeiten 124
3.7.1 Problemstellung 124
3.7.2 Erwartungswert 127
3.7.3 Varianz und Schätzwert der Varianz 130

3.8 Kombination der A-posteriori-Schichtung mit anderen Hochrechnungsverfahren 131
3.8.1 A-posteriori-geschichtete Verhältnisschätzung 133
3.8.2 A-posteriori-geschichtete Differenzenschätzung 135
3.8.3 A-posteriori-geschichtete Regressionsschätzung 136

3.9 Anwendung 137

3.10 Anwendungsbeispiele 140
3.10.1 Schätzung des Viehbestandes 140
3.10.1.1 Einfache Hochrechnung 141
3.10.1.2 A-posteriori-Schichtung 141
3.10.1.2.1 Hochrechnung mit den "wahren" N_h-Werten 142
3.10.1.2.2 Hochrechnung mit geschätzten N_h-Werten 143
3.10.1.2.3 Hochrechnung mit den M_h-Werten einer korrelierten Variablen 145
3.10.1.3 Vergleichende Darstellung 146
3.10.2 Schätzung des Angabefehlers 148
3.10.2.1 Modell B: Auswahl proportional zum Viehbestand 152
3.10.2.2 Modell C: Auswahl proportional zur Anzahl der Betriebe 153
3.10.2.3 Vergleich beider Modelle 154

4 ABSCHLIESSENDE BEMERKUNGEN 156

5 LITERATURVERZEICHNIS 159

Notationen

Die am häufigsten auftretenden Notationen in dieser Arbeit sind:

y	untersuchte Variable
x	Variable, die die zusätzliche Information darstellt
$\bar{Y}$, $\bar{X}$	Mittelwerte in der Grundgesamtheit
$\bar{y}$, $\bar{x}$	Mittelwerte in der Stichprobe (Schätzwerte)
$\hat{Y}$, $\hat{X}$	Schätzwerte des Totalwertes: $\hat{Y}=N\bar{y}$ bzw. $\hat{X}=N\bar{x}$
N	Anzahl der Elemente in der Grundgesamtheit
n	Anzahl der Elemente in der Stichprobe
L	Anzahl der Schichten
h	Nummer der Schicht
N_h, n_h	Zahl der Schichtmerkmale in der Grundgesamtheit bzw. in der Stichprobe
$P_h=\frac{N_h}{N}$	Schichtanteilssatz in der Grundgesamtheit
$p_h=\frac{n_h}{n}$	Schichtanteilssatz in der Stichprobe
$\hat{P}_h=\frac{\hat{n}_h}{\hat{n}}$	Schätzwert für P_h
p_i	Auswahlwahrscheinlichkeit
$\sigma_y^2=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N}$	Varianz der Grundgesamtheit, analog σ_x^2
S^2	ein Mehrfaches der Varianz der Grundgesamtheit: $S^2=\frac{N}{N-1}\sigma^2$, somit
$S_y^2=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N-1}$	, analog S_x^2

Allgemein gilt $S_y^2 = S^2$, wenn die Varianz der untersuchten Variablen nicht extra gekennzeichnet werden soll.

$S_{xy} = \frac{\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})}{N-1}$ Kovarianz in der Grundgesamtheit (s.o.)

$s_y^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$ Schätzwert von S_y^2, analog s_x^2 und s_{xy}

$V(\bar{y}_{..})$, $V(\hat{Y}_{..})$ "wahre" Varianz eines Schätzwertes (Mittelwert, Totalwert)

$v(\bar{y}_{..})$, $v(\hat{Y}_{..})$ Schätzwert der Varianz

$\rho = \frac{\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})}{\sum_{i=1}^{N}(x_i-\bar{X})^2 \sum_{i=1}^{N}(y_i-\bar{Y})^2} = \frac{S_{xy}}{S_x S_y}$ Korrelationskoeffizient zwischen x und y in der Grundgesamtheit

$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}$ Schätzwert des Korrelationskoeffizienten ρ

$B = \frac{\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})}{\sum_{i=1}^{N}(x_i-\bar{X})^2}$ Regressionskoeffizient zwischen x und y in der Grundgesamtheit

$b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ Schätzwert des Regressionskoeffizienten B

1 Einleitung

> **"But it would be a Mistake to ignore those whose interest is more purely intellectual, who desire an insight into the peculiar problems of probable inference and the structure of empirical knowledge, who wish to get a fundamental acquaintance with one of the most fundamental of subjects, to see and understand fully the mathematical derivations underlying so much practical and scientific activity and perhaps to make their own contributions."** [1)]

Die meisten Aussagen, die wir machen, treffen wir auf Grund von wenigen Beobachtungen. So ist die Stichprobentheorie die Wissenschaft, die quantitative Studien betreibt über Umfang, Verhaltensweisen, Bestandteile und Unterschiede sowie Unterschiedsgründe (Tests) beim Auftreten von Ereignissen. Das Ziel einer solchen Untersuchung ist, nützliche Informationen zur richtigen Zeit mit angebbaren und akzeptablen Vertrauensgrenzen zu erhalten[2)]. Dazu lehrt uns die Theorie "a) was wir über das Material wissen müssen, um eine Stichprobe sachgemäß durchführen zu können und b) welche Genauigkeit der vorgeschlagene Stichprobenplan erwarten läßt"[3)]. Bei statistischen Aussagen, die mit Hilfe dieser Stichproben gewonnen werden, ist darauf zu achten, daß sie auf der Grundlage von "statistischen Gesetzen" gewonnen werden[4)]. Das heißt, sie gelten nur für Mengen und/oder Mittelwerte von Erscheinungen, nicht für die Einzelerscheinung. Gini nennt das das statistische Gesetz erster Art. Das statistische Gesetz zweiter Art be-

---

1) Hotelling, Harold (1960): S.11.

2) Über das verspätete Erscheinen von statistischen Ergebnissen wird von den Konsumenten der Statistik oft Klage geführt. So stand in der Süddeutschen Zeitung am 1.4.1968 ein Artikel mit der Überschrift: "Sand in der Zahlenmühle. Die statistischen Behörden brauchen zwei Monate, um zehn Daten zu addieren." Eine stichprobenweise Voraufbereitung könnte hier oft Abhilfe schaffen.

3) Deming, Edwards W. (1960): S.25.

4) Gini, Corrado (1958): S.57.

trifft die Wahrscheinlichkeiten, mit denen die Ereignisse auftreten. Diese Wahrscheinlichkeiten müssen bei der Auswahl sowie bei der Hochrechnung der Daten berücksichtigt werden, sonst erhält man verzerrte oder gar falsche Ergebnisse.

Obwohl die Wahrscheinlichkeitsrechnung und die Stichprobentheorie noch relativ junge Wissenschaften sind, sind ihre Methoden heute schon weit verbreitet und werden auf weitere Gebiete ausgedehnt. Die Anwendung dieser Methoden nahm nach früheren Anfängen in Rußland und USA vor allem im Zweiten Weltkrieg und danach einen großen Aufschwung. In Industriebetrieben ab einer gewissen Größe gibt es keinen Bereich mehr, in dem nicht Statistik und Wahrscheinlichkeitsrechnung angewandt würden[1]. Das gleiche gilt für die Wissenschaft, die praktisch in keinem ihrer Zweige ohne Stichprobenmethoden mehr auskommt[2]. Eine der interessantesten Entwicklungen in neuerer Zeit ist die Verbindung der Stichprobentheorie mit der Entscheidungstheorie, an der sich Abraham Wald kurz vor seinem Tode sehr interessiert gezeigt hat[3].

Doch bevor solche neueren Entwicklungen auf breiter Basis diskutiert werden, sollte die Stichprobentheorie in ein abgeklärteres Stadium eintreten. Die Stichprobentheorie leidet mehr als andere Bereiche der Statistik am Mangel eines einheitlichen "roten Fadens", auf dem die verschiedenen Themenkreise, aus denen sie sich zusammensetzt, aufgereiht werden können[4]. Die wichtigsten davon sind Zufallsstichproben, Schichtung, Klumpung, Stufung und Hochrechnungsverfahren. Das erstrebenswerte Ziel ist, eine geschlossene Theorie zu finden, die dann der Praxis anwendbare Formeln und Rezepte liefert.

---

1) Auf die Bedeutung der Statistik und Wahrscheinlichkeitsrechnung in den Betrieben weist Weinberg, Franz (1968): S.10-14 hin; ebenso Mertz, Friedrich Georg (1966), der in einem Artikel die Schätzung des Betriebslagers auf Stichprobenbasis behandelt.

2) Vgl. Hotelling, Harold (1960): S.11.

3) Mahalanobis, P.C. (1952): S.1.

4) S. Stuart, Alan (1964): S.773.

Dazu müssen aber zuerst die einzelnen Problemkreise, wie sie eben angeführt wurden, geklärt werden. Zur Klärung dieser Einzelprobleme soll hier ein Beitrag geleistet werden, und zwar hinsichtlich eines speziellen Hochrechnungsproblems.

Der Begriff Hochrechnung - ohne daß er näher erklärt würde - ist durch die sehr frühzeitige Vorausschätzung der Ergebnisse von politischen Wahlen sehr in Mode gekommen. Vor allem deshalb, weil die Ergebnisse zu diesem Zeitpunkt, kurz nach Schließung der Wahllokale, schon sehr zutreffend sind.

Bei der Hochrechnung handelt es sich um eine Verbesserung der Ergebnisse mit Hilfe zusätzlicher Informationen, nachdem die Auswahl der Stichprobe erfolgt ist und die Daten vorliegen. Das heißt, wenn in die Schätzfunktion zusätzliche Variable (zusätzliche Information) eingebaut werden, erhalten wir engere Vertrauensgrenzen, und deshalb genauere Ergebnisse. Diese zusätzlichen Variablen müssen natürlich für das Untersuchungsmerkmal einen Aussagewert haben, d.h., sie müssen mit ihm gut korreliert sein. So ist beim Beispiel der Wahlen das Untersuchungsmerkmal die Abgabe der Wählerstimmen für eine Partei. Aus früheren Wahlen kennt man nun die einer Wahlentscheidung korrelierten Merkmale, wie Geschlecht, Alter, Religionszugehörigkeit, ländliche oder städtische Struktur des Wahlgebiets. Liegen nun die Ergebnisse einiger Wahlkreise vor, so werden diese mit Hilfe dieser korrelierten Merkmale adjustiert und hochgerechnet zu einem Schätzwert für den gesamten Staat.

Der Beitrag hier befaßt sich mit einem speziellen Hochrechnungsproblem, der A - p o s t e r i o r i - S c h i c h t u n g , das heißt der Verbesserung der Genauigkeit einer Stichprobe durch den Gebrauch von Schichtgrößen, wenn die Stichprobe schon vorliegt. Man könnte das Verfahren auch eine geschichtete Stichprobe nennen, im Gegensatz zum - aus Theorie und Praxis bekannten - geschichteten Auswahlverfahren. In der Literatur werden die Begriffe nicht immer klar getrennt. Das geschichtete Auswahlverfahren nennen wir hier auch A-priori-Schichtung, um zu zeigen, daß die Schichtung vor der Stichprobenauswahl vorgenommen wurde, im Gegensatz zur A-posterio-

ri-Schichtung, bei der die Schichtung nach der Auswahl vorgenommen wird.

Im nächsten Abschnitt sollen die Voraussetzungen zur A-posteriori-Schichtung als Schätz- bzw. Hochrechnungsverfahren behandelt werden. Unter den Voraussetzungen werden auch kurz die gebräuchlichsten Hochrechnungsarten in ihrer Problematik gestreift, weil diese im Laufe der Arbeit zu Wirksamkeitsvergleichen herangezogen und in kombinierten Schätzverfahren mit der A-posteriori-Schichtung behandelt werden.

2 Voraussetzungen

2.1 Schätzprobleme im Allgemeinen

2.1.1 Geschichtlicher Rahmen

Der Schluß vom Teil aufs Ganze als Methode zur Ermittlung neuer Kenntnisse - ob im täglichen Leben oder in der Wissenschaft, ja selbst in der Metaphysik - wird wohl seit Bestehen der Menschheit mehr oder weniger bewußt angewandt. Methodisch wurde eine Art von Stichproben zur Erfassung der Bevölkerung schon im Mittelalter angewandt, indem von der Zahl der Feuerstellen auf die Bevölkerungsgesamtzahl hochgerechnet wurde.[1)] Doch erst im 17. und 18. Jahrhundert versuchte man - zunächst für das Gebiet der Bevölkerungsstatistik - Gesetzmäßigkeiten festzustellen. Hier sind an erster Stelle die Namen Graunt, Halley, Petty und Süßmilch zu nennen. Einen recht modern anmutenden Stichprobenplan zur Bevölkerungsschätzung entwarf Moheau im Jahre 1778, indem er Frankreich in Distrikte aufteilte und von diesen einen Teil repräsentativ, aber einige

---

1) S. Strecker, Heinrich (1957): S.9.

auch mit Hilfe des Zufalls auswählte und daraus einen Schätzwert berechnete.[1] Völlig neue Möglichkeiten für die Statistik, insbesondere für die Stichprobentheorie, brachte die Wahrscheinlichkeitsrechnung, deren Grundlagen auf den Arbeiten von Fermat, Pascal und Bernoulli fußen. Als erster führte dann Laplace (1802) auf der Grundlage der Wahrscheinlichkeitstheorie eine Bevölkerungsschätzung in Frankreich durch.[2]

Von der reinen Bevölkerungszählung weg und schon als Ansatz zu einer Entwicklung der Statistik als methodische Hilfswissenschaft ohne speziellen Anwendungsbereich, ist die Arbeit von Quetelet "Über den Menschen und die Entwicklung seiner Fähigkeiten oder Versuch einer Physik der Gesellschaft"(1835) anzusehen. In dieser Richtung die statistische Methodenlehre als methodische Hilfswissenschaft fortgeführt zu haben, ist vor allem das Verdienst der Engländer Galton, Pearson und R.A. Fisher, die alle das Problem von der Naturwissenschaft (Biologie) her anfaßten.[3]

In der amtlichen Statistik setzte sich das Stichprobenverfahren nur langsam durch. Hierbei sind vor allem in Norwegen die Arbeit von Kiär[4], in Rußland die von Tschuprow[5] und in England die von Bowley[6] zu nennen, wobei letzterer die Anwendung mathematischer Methoden für unumgänglich hielt und in dieser Richtung die statistische Methodenlehre weiter festigte. In Deutschland, wo die Statistik durch die historische Schule der Nationalökonomie beeinflußt wurde, herrschte große Abneigung gegenüber der Anwendung mathematischer Methoden, die erst allmählich überwunden wird.[7] Doch auch hier haben

---

1) S. Szulc, Stefan (1965): S.568.

2) Vgl. Stephan, Frederick F. (1948): S.15.

3) Vgl. Kellerer, Hans (1960): S.278.

4) Kiär, A.N. (1899): S.1-22.

5) S. Strecker, Heinrich (1957): S.11.

6) Bowley, A.L. (1906): S.540-548.

7) Hier trifft gut die Bemerkung von Bowley, A.L. (1906): S.545 zu: "The use of mathematical reasoning in statistics is very imperfectly understood, partly because the passage from numbers to symbols and back to numbers suggests an air of mystery, or even of prestidigation, to the unmathematical mind ...."

vereinzelt Männer wie Mayet[1] und Schott[2] schon früh Stichproben in der amtlichen Statistik durchgeführt. Stark vorangetrieben wurde die Stichprobenmethode durch die Arbeit des Internationalen Statistischen Instituts, nach deren Richtlinien O. Anderson 1926 eine stichprobenweise Aufbereitung der landwirtschaftlichen Betriebszählung in Bulgarien durchführte.[3]

In der Folgezeit bis heute setzte ein großer Aufschwung in der Theorie wie in der Anwendung der Stichproben ein. Die systematische Entwicklung der Theorie für die Anwendung von Stichprobenerhebungen beginnt nach Sukhatme mit Neyman (1934).[4] Beträchtliche Fortschritte wurden hier im Jahrzehnt 1940-50 gemacht[5]. Sie sind eng verbunden mit den Namen Sukhatme, Cochran, Stephan, Hansen, Hurwitz, Madow, Mahalanobis, Dalenius, Zarkovich und vielen anderen. Behandelt wurden Probleme der Auswahl, der Schichtung und der Hochrechnung, wo bei letzterer mit Hilfe zusätzlicher Informationen neue Hochrechnungsmethoden entwickelt wurden, und eine davon hier behandelt werden soll. Zur Zeit haben wir eine "natürliche Fortentwicklung" dieser in den 30iger und 40iger Jahren entwickelten Probleme; man sucht neue Varianzmodelle bei Zufallsstichproben, wirkungsvolle Anwendungsverfahren für bestimmte Gebiete und vereinfachte Verfahren bei komplizierten Stichprobenmodellen.[6] Ein großer Förderer der Statistik im allgemeinen, sowie der Stichprobenmethoden im besonderen - vor allem wegen ihrer Anwendbarkeit in Entwicklungsländern - ist das Statistische Amt der Vereinten Nationen.[7] Heute fin-

---

1) Mayet, P. (1905): S.258-276.

2) Schott, Sigmund (1917): S.279-314.

3) Anderson, Oskar (1949).

4) Sukhatme, P.V. (1966): S.367.

5) Vgl. Murthy, M.N. (1963): S.737.

6) Vgl. Dalenius, Tore (1962): S.329.

7) In den Publikationen der Vereinten Nationen finden wir auch eine gute Geschichte der Anwendung der Statistik in den einzelnen Ländern. So in: United Nations (Hrsg.)(1953) und (1955).

den wir Stichprobenmethoden in der Physik, Astronomie, Biometrie, Soziometrie usw.. So haben sich ganz neue Wissenschaftszweige entwickelt, wie in jüngster Zeit die Jurimetrie.[1)]

2.1.2 Grundlagen der Schätztheorie

2.1.2.1 Das Schätzen von Parametern

Zur statistischen Analyse gehört das Schätzen eines Parameters der Grundgesamtheit aus Stichprobendaten. Dabei nennen wir jene Werte Parameter, wie z.B. den Mittelwert oder die Streuung, die die Eigenschaft haben, eine Grundgesamtheit von Elementen unzweideutig zu kennzeichnen und zu beschreiben,[2)] und diese beiden Parameter besitzen diese Eigenschaft, soweit - bei Unkenntnis der Verteilung der Variablen - der zentrale Grenzwertsatz seine Gültigkeit hat. Diese wichtige Voraussetzung wird in der Literatur zwar meist angenommen, aber auch meist stillschweigend. Auf der Basis einer relativ kleinen Stichprobe also machen die Statistiker Aussagen über die unbekannten Parameter einer Grundgesamtheit. Diese sind Konstanten, solange sich die Grundgesamtheit im Laufe der Zeit nicht ändert. Es leuchtet sofort ein, daß ein Schätzwert nur falsch oder richtig sein kann - in den angegebenen Grenzen; er kann nicht beinahe richtig sein. Die Aussage, "das Durchschnittsalter der Hausbesitzer liegt zwischen 40 und 45 Jahren", ist entweder richtig oder falsch.[3)] Die Berechtigung und Begründung dieses Vorgangs - von den vorhandenen und analysierten Stichprobendaten auf die Grundgesamtheit zu schließen - soll uns in den folgenden Abschnitten kurz beschäftigen.

---

1) S. Solomon, Herbert (1966).
2) Vgl. Mack, C. (1966): S.63.
3) Vgl. Richmond, Samuel B. (1964): S.158f..

2.1.2.2 Statistische Schlußweise

Aus den von den Stichprobendaten gelieferten Informationen werden Schlüsse gezogen und Entscheidungen getroffen, die sich auf die untersuchte oder "zukünftige" Grundgesamtheit beziehen. Dieser Vorgang wird als statistische Schlußweise (Inferenz) bezeichnet. Sie kann auch kurz folgendermaßen definiert werden: "Statistische Inferenz ist die Überwindung der Ungewißheit durch induktive Schlüsse, die ihre Basis in empirischen Beobachtungen haben"[1]. Im folgenden Schaubild soll dieser Vorgang schematisch dargestellt werden[2].

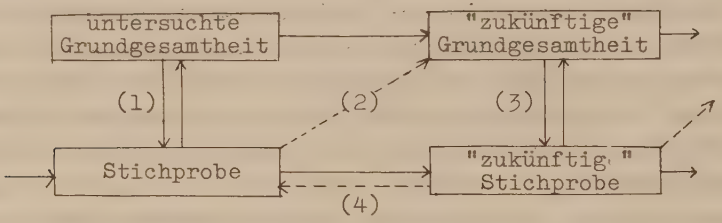


Abb.1 Schematische Darstellung des statistischen Schließens

(1) Wir ziehen eine Stichprobe und charakterisieren damit die untersuchte Grundgesamtheit.
(2) Wir schließen auf die "zukünftige" Grundgesamtheit und treffen darüber Entscheidungen.
(3) Wir ziehen in der Zukunft aus dieser Grundgesamtheit eine neue "zukünftige" Stichprobe. Es folgen wieder die Vorgänge (1) und (2).
(4) Wir schließen auf die Qualität der gezogenen ersten Stichprobe und verbessern eventuell den Stichprobenplan.

---

1) Menges, Günter (1968): S.266. Menges bringt an dieser Stelle, S.266/267, auch breite Ausführungen über den Inferenzbegriff, der neu in der deutschen Sprache ist.

2) S. Springer, C.H./Herlihy, R.E./Mall, R.T./Beggs, R.T. (1966): S.205/206.

Traditionell werden die Probleme des statistischen Schließens, wobei über die grundlegenden Prinzipien ein beträchtlicher Streit entstand[1], in folgende Problemkreise eingeteilt[2]:

Schätzprobleme
Probleme der Hypothesenprüfung
Entscheidungstheorie

Dabei ist die Entscheidungstheorie jüngeren Datums als die beiden anderen Problemkreise und ist grundlegend von Abraham Wald 1950 entwickelt worden[3].

In dieser Arbeit sollen uns ausschließlich Schätzprobleme interessieren und unter diesen nur Punktschätzungen. Bei Punktschätzungen schätzen wir, wie der Name schon sagt, einzelne Zahlen, einzelne Punkte, so den Mittelwert oder die Streuung; anders dagegen bei Intervallschätzungen, wo wir das Konfidenzintervall angeben, das mit einer vorgegebenen Wahrscheinlichkeit den Parameter überdeckt. Dann liegt der unbekannte gesuchte Parameter Θ mit einer bestimmten Wahrscheinlichkeit zwischen den Grenzen u_1 und u_2 (diese Grenzen haben wir aus einer Stichprobe geschätzt):

$$P(u_1 \leqq \Theta \leqq u_2) = \gamma$$

Es ist oft eine Frage der Zweckmäßigkeit, ob man Punktschätzungen oder Intervallschätzungen anwendet.

Die Parameter bei Punktschätzungen werden mit Hilfe von Schätzfunktionen (oder einfach Schätzungen) geschätzt. Dadurch erhält man den Schätzwert, der eine Konstante ist. Der Unterschied zwischen einer Schätzung (Schätzfunktion) und einem Schätzwert ist der gleiche wie zwischen einer Funktion f(x) und dem Funktionswert f(a). Allgemein gilt[4]: Θ sei der

---

1) Ausführliche Darstellung der Inferenzprobleme und -theorien sind zu finden bei Weibull, Christer (1960), aber auch bei Menges, Günter (1968): S.266-290.

2) Vgl. Bulmer, M.G. (1965): S.144ff..
Freund, John E. (1962): S.209ff..

3) Vgl. Bulmer, M.G. (1965): S.159.

4) Vgl. Wine, Lowell R. (1964): S.170/171.

unbekannte Parameter einer Verteilung mit der Dichtefunktion $f(x)$. $x_1, x_2, \ldots, x_n$ sei eine Zufallsstichprobe der Größe n aus der untersuchten Grundgesamtheit. Bevor ein bestimmtes Element in die Stichprobe fällt, kann die Zufallsvariable x jeden Wert annehmen. Wenn wir i=1,2,...,n Stichproben ziehen, ist

$$t_n = t(x_1, \ldots, x_n)$$

die Schätzfunktion von O, und wenn eine bestimmte Stichprobe gezogen ist, der Schätzwert für Θ. So ist $\bar{x}$ als Schätzung für den Mittelwert μ eine Zufallsvariable, aber $\bar{x} = 8{,}4$ ist als Schätzwert eine Konstante. Wie leicht einsichtig ist, ist unser Problem nicht das Studium der Schätzwerte, sondern das Studium der Schätzfunktionen.

Nun liefern uns unterschiedliche Schätzfunktionen unterschiedliche Schätzungen[1]. Nehmen wir an - die Gültigkeit des zentralen Grenzwertsatzes vorausgesetzt -, wir hätten 3 Schätzfunktionen $t_1(x_1,..,x_n)$, $t_2(x_1,...,x_n)$ und $t_3(x_1,...,x_n)$ zur Schätzung des Parameters Θ. Dann können wir z.B. folgende Schätzverteilung der Stichprobenwerte erhalten:

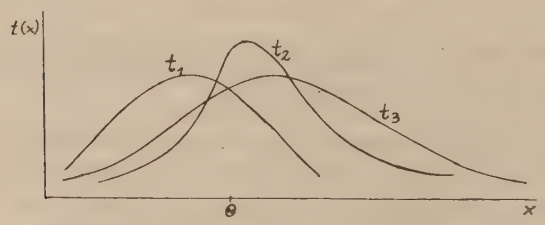


Abb.2 Verteilungen von Stichprobenschätzungen

Welches nun die Eigenschaften guter Schätzfunktionen sind, beschäftigt uns im nächsten Abschnitt.

---

[1]) Vgl. Bowker, A.H./Liebermann, G.J. (1961): S.211ff..

2.1.2.3 Eigenschaften guter Schätzfunktionen

Wie wir gesehen haben, können verschiedene Schätzfunktionen unterschiedlich gute Schätzwerte liefern. Auf Grund welcher Eigenschaften können wir nun entscheiden, welches die Schätzfunktion ist, die uns in einer gegebenen Situation in kurzer Zeit, bei geringem Risiko und geringen Kosten die beste Information durch die Schätzwerte liefert?

Die vier wichtigsten statistischen Eigenschaften dafür sind: Unverzerrtheit bzw. asymptotische Unverzerrtheit, Konsistenz, Effizienz und Suffizienz. Um die Aufstellung dieser Kriterien hat sich der englische Statistiker R.A. Fisher sehr verdient gemacht. Diese vier Kriterien, die die Grundlage für gute Schätzungen bilden, wollen wir im folgenden kurz beschreiben.[1)]

2.1.2.3.1 Verzerrte und unverzerrte Schätzung

Das wichtigste - weil in der praktischen Stichprobenplanung am meisten beachtete - Kriterium ist die Unverzerrtheit oder Erwartungstreue der Schätzung. Generell sollen die Schätzmethoden unverzerrte Schätzwerte liefern, d.h., der Durchschnitt aller $\binom{N}{n}$ möglichen Stichprobenmittelwerte $\bar{x}$ muß gleich dem wahren Wert $\bar{X}$ der Grundgesamtheit sein:[2)]

$$E(\bar{x}) = E\left(\frac{x_1+x_2+\ldots+x_n}{n}\right) = \bar{X}$$

Ebenso soll $s^2 = \frac{\sum_{1}^{n} (x_i-\bar{x})^2}{n-1}$ ein unverzerrter Schätzwert für

---

1) Eine ausführliche Behandlung der mathematischen Form dieser Kriterien, auf die wir nicht eingehen, ist bei Schmetterer, Leopold (1966): S.324ff. zu finden.

2) Die Herleitung des folgenden kann man praktisch in jedem Stichprobenlehrbuch finden, u.a. Cochran, William G. (1964): S.20ff.; Kapur, J.N./Saxena, H.C. (1963): S.274f.; Wine, Lowell R. (1964): S.172ff..

$$S^2 = \frac{\sum_{1}^{N} (x_i - \bar{X})^2}{N-1} \quad \text{sein}^{1)}.$$

Der Erwartungswert wird:

$$E(s^2) = \frac{1}{n-1} E\left[\Sigma(x_i - \bar{X} - \bar{x} + \bar{X})^2\right]$$

$$= \frac{1}{n-1} E\left[\Sigma(x_i - \bar{X})^2 - n(\bar{x} - \bar{X})^2\right]$$

Nun ist

$$E\left[\Sigma(x_i - \bar{X})^2\right] = n\sigma^2 = \frac{n(N-1)}{N} S^2$$

und

$$E\left[n(\bar{x} - \bar{X})^2\right] = \frac{N-n}{N} S^2$$

Also

$$E(s^2) = \frac{1}{n-1}\left[\frac{n(N-1)}{N} S^2 - \frac{N-n}{N} S^2\right] = S^2$$

Aus dem Gesagten erhalten wir die Definition der Verzerrung, oder wie es üblicher ist, des "Bias" (das Wort stammt aus der angelsächsischen Literatur und hat sich international eingebürgert):

$$B = E(\bar{x}) - \bar{X}$$

Bei unverzerrten Schätzmethoden ist der Bias definitionsgemäß gleich O. Der Vorteil dieser Methoden liegt auf der Hand; die Ungenauigkeit der Schäzuwerte kann allein durch die Varianz angegeben werden, ohne auf den störenden Einfluß eines eventuellen Bias einzugehen.

Es hat sich nun gezeigt, daß Schätzungen mit zusätzlichen In-

---

1) Gewöhnlich werden in er Stichprobenliteratur die Formeln mit S^2 (im Nenner (N-1)) anstatt mit σ^2 (im Nenner N) angewandt. Es gilt:

$$S^2 = \frac{N}{N-1} \sigma^2 .$$

Die Varianzformeln erhalten dadurch einfachere Ausdrücke (Vgl. Cochran, William G. (1964): S.22).

formationen (gebundene Hochrechnung) bessere Schätzwerte liefern können, d.h., die Genauigkeit (precision) ist größer, weil sie kleinere Standardabweichungen liefern als die einfache Hochrechnung. Diese Tatsache läßt sich einfach aus der Anwendung der zusätzlichen Information erklären. Diese Schätzmethoden sind jedoch nicht immer erwartungstreu. Es kann ein Bias auftreten; $B = E(\bar{x} - \bar{X})$ wird ungleich 0. Der Durchschnitt aller möglichen Stichprobenwerte ist nicht gleich dem wahren Wert, wir erhalten also ein verzerrtes Ergebnis. Aber "der Schönheitsfehler, daß die Anlage des Stichprobenverfahrens oder die Berechnung von Schätzwerten 'verzerrt' ist, kann evtl. dadurch mehr als ausgeglichen werden, daß die Streuung wesentlich herabgedrückt wird.... In der Wirklichkeit ist die Verzerrung meist recht unbedeutend."[1] Ferner nimmt der Bias bei konsistenten Schätzwerten rasch ab und strebt gegen 0, wie wir in unseren Ausführungen an späterer Stelle noch sehen werden.

Zur Erläuterung dieses Tatbestandes der größeren Genauigkeit oder besser der größeren Wirksamkeit (s. Abschnitt 2.1.2.3.3) soll folgendes Schaubild dienen:[2]

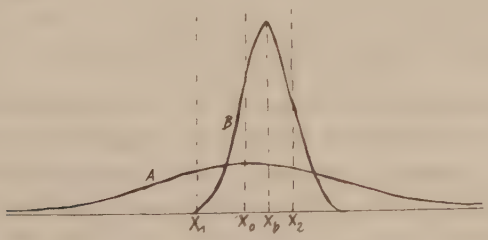


Abb.3 Zwei Verteilungen von Stichprobenschätzungen: unverzerrte Schätzungen in Verteilung A; verzerrte Schätzungen mit kleinerer Standardabweichung in B.

Quelle: Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a): S.35.

---

1) Kellerer, Hans (1963): S.181.

2) Entnommen aus Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a): S.35.

Im Schaubild stellt A die Verteilung von Stichprobenschätzungen dar, deren Erwartungswert gleich dem wahren Wert X_0 ist. Aber relativ wenige dieser Stichprobenschätzungen liegen nahe bei X_0, das soll hier heißen: fallen in das Intervall $(X_1 X_2)$. Betrachten wir dagegen die Verteilung von Stichprobenschätzungen wie B sie darstellt, und deren Erwartungswert X_b nicht dem wahren Wert X_0 entspricht, so liegen hier wesentlich mehr Schätzungen nahe bei dem wahren Wert X_0, obwohl diese Schätzmethode verzerrt ist. In solch einem Fall besteht eine größere Chance, mit Hilfe von verzerrten Schätzmethoden (Fall B) "gute" Stichproben zu erhalten, als bei unverzerrten Methoden (Fall A), in denen viele der Stichproben große Stichprobenfehler aufweisen.

Anmerkungen zum Bias:

(1) Der Ausdruck "Unverzerrtheit" ist viel mehr eine Eigenschaft der Schätzmethode als des Schätzwertes. Die Feststellung, daß "$\bar{y}_n$ eine unverzerrte Schätzung von $\bar{y}_N$ ist", ist eine wirkliche Aussage.[1] Dagegen sagt es strenggenommen wenig aus, daß "168,5 ein unverzerrter Schätzwert von $\bar{y}_N$ ist". Letzteres wird zwar allgemein praktiziert und bedeutet meist implizit, daß es sich um eine unverzerrte Schätzmethode handelt. Das darf aber nicht dazu führen, daß vergessen wird, daß ein einzelner Schätzwert beträchtlich vom Wert der Grundgesamtheit abweichen kann, obwohl er "unverzerrt" ist.

(2) "Der oben beschriebene Bias ist ein mathematischer, das heißt, er ist der Schätzung inhärent. Die Größe dieses 'internen' Bias ist bekannt und kann durch entsprechende Korrekturen beseitigt werden. Eine Schätzung kann aber auch mit einem nichtmathematischen, 'externen' Bias behaftet sein, der durch experimentelle Meß- oder Beurteilungsfehler oder durch nichtzufälliges Stichprobensammeln oder durch beides zusammen verursacht wurde. Dieser Bias ist nicht so harmlos wie der mathematische, weil er sich in den seltensten Fällen in seinem Ausmaß und seiner Richtung erkennen und damit eliminieren läßt. Externer Bias läßt sich nur durch sorgfältige Pla-

---

[1]) S. Sampford, M.R. (1962): S.23.

nung der Experimente (experimental design) verhüten."[1)]

(3) In der vorliegenden Arbeit haben wir es hauptsächlich mit "biasfreien" oder "asymptotisch biasfreien" Schätzungen zu tun. Asymptotisch biasfrei heißt: der Bias ist vom Stichprobenumfang abhängig. Er ist dann in der Regel bei kleinen Stichproben größer und tendiert gegen 0, wenn der Umfang der Stichprobe gegen jenen der Grundgesamtheit strebt.[2)]

2.1.2.3.2 Konsistenz

Das Kriterium der Unverzerrtheit ist ein sehr allgemeines, es besagt lediglich, daß im Durchschnitt alle Schätzwerte gleich dem wahren Wert sind. Aber wie nahe die Schätzwerte dem wahren Wert zu liegen kommen, darüber sagt es nichts aus. Diese Forderung, daß die Schätzwerte sich mit großer Wahrscheinlichkeit um den wahren Wert eng gruppieren, erfüllt das Kriterium der Konsistenz.

Man definiert: Eine Schätzung t_n, die aus einer Zufallsstichprobe von n Werten errechnet wurde, ist eine konsistente Schätzung für Θ, wenn - bei zunehmendem Stichprobenumfang, d.h., der Umfang der Stichprobe tendiert gegen den der Grundgesamtheit - t_n mit Wahrscheinlichkeit gegen Θ konvergiert.[3)] Somit neigen alle "konsistenten Statistiken ... dazu, mit zunehmender Größe der Beobachtungsreihe den korrekten Werten näher und näher zu kommen".[4)]

Mathematisch ausgedrückt heißt das bei einer unendlichen Grundgesamtheit, bzw. im Fall mit Zurücklegen[5)]: Bleibt ein Parameter Θ mit wachsendem Stichprobenumfang n konstant, so gilt: Die Wahrscheinlichkeit, daß der Absolutwert der Diffe-

---

1) Geigy, J.R. A.G.(Hrsg.)(1960): S.153, ebenso: Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a): S.39.

2) Vgl. Geigy, J.R. A.G.(Hrsg.)(1960): S.153.

3) Thionet, M.P. (1953b): S.79.

4) Fisher, Ronald A. (1956): S.11.

5) Aus Geigy, J.R. A.G.(Hrsg.)(1960): S.154.

renz aus Schätzung t_n und ihrem Erwartungswert Θ eine beliebig kleine Zahl ε nicht übersteigt, strebt gegen 1, wenn der Stichprobenumfang n gegen Unendlich strebt. In Symbolen ausgedrückt:

$$\lim_{n \to \infty} P(|t_n - \Theta| < \varepsilon) = 1 \qquad (\varepsilon > 0)$$

Für Parameter, die sich mit wachsendem Stichprobenumfang n proportional zu n, n^2 usw. vergrößern, gilt das oben Gesagte auch, wenn der Absolutwert der Differenz aus Schätzung minus Erwartungswert durch n, n^2 usw. dividiert wird.

Wir sehen also: Schätzungen, die dem Gesetz der großen Zahlen genügen - mit nichts anderem haben wir es hier zu tun -, nennt man konsistent.[1)]

Anmerkungen:[2)]

(1) Die Eigenschaft der Konsistenz sichert, daß die Differenz zwischen t_n und Θ mit Wahrscheinlichkeit immer kleiner wird, wenn n sich unendlich vergrößert.

(2) Diese Eigenschaft ist eine "Grenz"-Eigenschaft, d.h., sie ist auf das Verhalten von t_n bei zunehmendem Stichprobenumfang zurückzuführen.

(3) Eine konsistente Schätzung ist im Grenzwert unverzerrt, aber eine unverzerrte Schätzung muß nicht konsistent sein.

Zwei Beispiele für konsistente Schätzwerte wollen wir noch anführen. Bei normalverteilter Grundgesamtheit, bei der Median, häufigster Wert (mode) und arithmetisches Mittel zusammenfallen, ist $\bar{x}$ für alle drei ein konsistenter Schätzwert.[3)] Da $\bar{x}$ ein unverzerrter Schätzwert des Mittelwertes der Grundgesamtheit mit Varianz σ^2/n ist, werden die Schätzwerte $\bar{x}$ der Stichprobenverteilung bei zunehmendem n immer enger um den Mittelwert der Grundgesamtheit zu liegen kommen.

Ebenso ist s^2 ein konsistenter Schätzwert von σ^2 bzw. S^2. Das läßt sich folgendermaßen zeigen: Es gilt folgendes Theorem[4)]:

---

1) Geigy, J.R. A.G.(Hrsg.)(1960): S.154.
2) S. Kapur, J.N./Saxena, H.C. (1963): S.392.
3) S. Wine, Lowell R. (1964): S.171/172.
4) Freund, John E. (1962): S.195.

Wenn s^2 die Varianz einer Zufallsstichprobe vom Umfang n einer normalverteilten Grundgesamtheit und erwartungstreu ist, dann hat die Variable $(n-1)s^2/\sigma^2$ eine χ^2-Verteilung mit n-1 Freiheitsgraden. Für die Varianz erhalten wir: 2(n-1).

Somit

$$\mathrm{Var}\left[(n-1)s^2/\sigma^2\right] = 2(n-1)$$

also

$$\mathrm{Var}(s^2) = \frac{2(n-1)\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{n-1}$$

Mit zunehmendem n scharen sich die Stichprobenvarianzen immer enger um die Varianz der Grundgesamtheit.

2.1.2.3.3 Effizienz (Wirksamkeit)

In der Stichprobentheorie errechnen wir die Schätzungen aus zufälligen Ereignissen. Diese schwanken deshalb von Stichprobe zu Stichprobe innerhalb eines gewissen, vom Stichprobenumfang abhängigen Bereichs um ihren Erwartungswert. Die Genauigkeit der Schätzung hängt natürlich von der Breite des Streuungsbereiches, d.h. von der Standardabweichung bzw. der Varianz ab. Somit können wir "jetzt aus der Gesamtheit der einheitlichen Statistiken eine Gruppe absondern, die besonders wertvoll ist und die man als .. effiziente Statistiken bezeichnet".[1)]

Angenommen, wir haben für große Stichproben zwei konsistente Schätzungen t_n und t_n', die beide um den wahren Wert des Parameters Θ mit der Varianz σ^2/n bzw. $\sigma^{2'}/n$ normal verteilt sind. Auf Grund des zentralen Grenzwertsatzes wird dies gewöhnlich der Fall sein. Wir bezeichnen dann t_n als eine "wirksamere" Schätzung als t_n', wenn $\sigma^2 < \sigma^{2'}$. So liefert die Schätzung aus einer normalverteilten Grundgesamtheit mit Hilfe des arithmetischen Mittels effizientere Schätzwerte als jene mit Hilfe des Medians; denn die Varianz des Stichprobenmittelwertes ist

---

1) Fisher, Ronald A. (1956): S.13. In der Übersetzung steht statt effizient wirkungsgesetzmäßig. Diese Übersetzung scheint etwas geschwollen.

bekanntermaßen σ^2/n und die des Medians $\pi\sigma^2/2n$.[1] Die relative Wirksamkeit des arithmetischen Mittels ist:

$$\frac{\frac{\pi\sigma^2}{2n}}{\frac{\sigma^2}{n}} = \frac{\pi}{2} = 1{,}57$$

Mit anderen Worten, das arithmetische Mittel ist um 57% wirksamer als der Median (Zentralwert). Oder man kann auch sagen, für die gleiche Wirksamkeit braucht man beim Median n=100 Stichprobeneinheiten, dagegen beim arithmetischen Mittel nur n=64 Stichprobeneinheiten[2]. Die Schätzung mit der kleineren Varianz gruppiert sich enger um den Mittelwert und wird im Durchschnitt weniger davon abweichen als die Schätzung mit der größeren Varianz (vgl. Abbild 3). Diese wird deshalb vernünftigerweise auch als wirksamer betrachtet. Finden wir gar eine konsistente Schätzung t_n, deren Varianz kleiner ist als die aller anderen Schätzungen, dann wird t_n als die wirksamste Schätzung bezeichnet. Die "Wirksamkeit" der anderen Schätzungen wird dann definiert als das Verhältnis der Varianz der gegebenen Schätzung zur Varianz der wirksamsten Schätzung.[3]

Wir können auch jene biasfreie bzw. asymptotisch biasfreie Schätzung als wirksamste bzw. asymptotisch wirksamste definieren, die der Ungleichung von Rao-Cramér genügt[4]. Eine solche Schätzung bezeichnen wir als 100% wirksam[5]. Jedoch gibt es solche 100% wirksame Schätzungen nur in Ausnahmefällen, dagegen gibt es asymptotisch wirksame Schätzfunktionen unter allgemeinen Voraussetzungen[6]. Im allgemeinen werden wir es in der praktischen Stichprobenanwendung mit relativ wirksamen

---

1) Vgl. Cramér, Harald (1963): S.309.
2) Vgl. Freund, John E. (1962): S.219.
3) Vgl. Kapur, J.N./Saxena, H.C. (1963): S.395.
4) Cramér, Harald (1963): S.477ff..
5) S. Geigy, J.R. A.G.(Hrsg.)(1960): S.154.
6) Schmetterer, Leopold (1966): S.409.

Schätzungen zu tun haben, wie das obige Beispiel von arithmetischem Mittel und Median gezeigt hat.

Für die Wirksamkeit einer Schätzung ist also der Stichprobenumfang und die Schätzmethode ausschlaggebend. In der Praxis beeinflussen nun neben mathematischen (Schätzmethode) mit Recht auch praktische (große Kosten bei großem Stichprobenumfang) Anforderungen unsere Überlegungen. Deshalb sollen hier einige "Faustregeln" angeführt werden.

"Man verwende die wirksamste aller Schätzungen eines Parameters.
- wenn die Versuche im Vergleich zur numerischen Auswertung kostspielig sind,
- wenn die Versuche nicht mehr wiederholt werden können,
- wenn χ^2-Teste geplant sind,
- wenn das Resultat möglichst präzis und informativ sein soll,
- wenn von anderer Seite üblicherweise wirksamste Schätzungen für gleiche Untersuchungen verwendet wurden (man verdirbt sich dann nicht die Möglichkeit gegenseitiger Vergleiche und Signifikanzteste).

Wenn keine Überlegungen dagegen sprechen, verwende man weniger wirksame Schnellschätzungen,
- wenn die numerische Auswertung teurer ist als die Versuche,
- wenn die Präzision der Schätzung für den vorgesehenen Zweck genügt,
- wenn die Versuche zur Vororientierung schnell überprüft werden sollen,
- für Routineuntersuchungen,
- zur Nachkontrolle wirksamerer Schätzungen, deren Berechnung viele Fehlermöglichkeiten aufweist."[1)]

2.1.2.3.4 Suffizienz (erschöpfende Schätzungen)

Für praktische Zwecke genügt meist das Kriterium der Effizienz. Das Wirksamkeitskriterium ist aber in einem gewissen Maße un-

---

1) Aus Geigy, J.R. A.G.(Hrsg.)(1960): S.154/155.

vollständig, denn bei großen Beobachtungsreihen neigen - trotz verschiedener Kalkulationsmethoden - wirksame Schätzungen dazu, gleichwertig zu sein[1]. Das vollständige Kriterium einer Schätzfunktion ist deshalb, sie so zu wählen, daß sie die "Gesamtheit der relevanten Informationen durch die Stichprobe liefert"[2]. Die gebräuchlichsten Interpretationen dieses sogenannten Kriteriums der Suffizienz sind[3]:

- die klassische, nämlich, daß wiederholte Stichproben dieselbe Verteilung erzeugen;
- die Bayes'sche, nämlich, daß die A-priori-Verteilung der Parameter jener der gefundenen Parameter aus der A-posteriori-Verteilung entspricht;
- die ökonomische, nämlich, daß ein Entscheidungsproblem und eine Entscheidungsprozedur durch einen experimentell gegebenen Schätzwert gelöst werden können.

Wie wir oben gesehen haben, liefert eine suffiziente Schätzung alle Informationen, die überhaupt möglich sind. Aber unglücklicherweise sind erschöpfende Schätzungen in der Praxis eher die Ausnahme als die Regel[4]. Darüber hinaus wird das Suffizienzkriterium meist in abstrakten Termini beschrieben, so daß es "vom Standpunkt der Stichprobentheorie aus weder notwendig noch wichtig" ist[5].

2.2 Stichproben

2.2.1 Der Stichprobenplan

Obwohl in der Literatur viel über Vor- und Nachteile von Stich-

---

1) Vgl. Fisher, Ronald A. (1956): S.14/15.
2) Fisher, Ronald A. (1950): S.316.
3) Vgl. Le Cam, L. (1964): S.1419ff..
4) S. Kapur, J.N./Saxena, H.C. (1963): S.398.
5) Pathak, P.K. (1964): S.795.

proben zu finden ist[1], scheint es doch einleuchtend, daß es keine Nachteile bei Stichproben gibt, sondern daß es sich nur darum handeln kann, ob es zweckmäßig ist, solche anzuwenden. Ein Stichprobenmodell ist klar und eindeutig definiert. Was meist als Nachteil angegeben wird, z.B. keine Einzelinformation und/oder keine sehr detaillierten Ergebnisse, sind Anforderungen, denen das Modell keine Rechnung trägt, weil es für diese Fälle nicht konstruiert worden ist. Auch bei der Anwendung von Stichproben gilt, was Pareto[2] für logisches Verhalten fordert, nämlich: es muß ein Ziel oder Zweck vorhanden sein; dieses Ziel muß erreichbar sein; die unternommenen Schritte oder angewandten Mittel zur Erreichung des Ziels müssen angemessen sein.

Die immer wieder angeführten Gründe für Stichprobenuntersuchungen sind: Zeit- und Kostenersparnis, freie Wahl der Konzeption (Stichprobenmodell), weiter Anwendungsbereich der Stichproben, der Stichprobenfehler kann angegeben und den Erfordernissen angepaßt werden, gute Anwendung der Stichproben für analytische Studien.

Voraussetzung für jede gute Stichprobe ist der Stichprobenplan, der die angewandten Methoden klar aufzeigt und einer Nachprüfung standhält. Damit ist die Durchführung "wilder" Stichproben - die meist auch spektakuläre Ergebnisse zeitigen[3] - unmöglich gemacht. Das Ziel einer Stichprobe liegt nicht in ihr selbst, sondern sie soll nützliche Informationen zur richtigen Zeit liefern. Und um diese Information zu erhalten, gilt, daß der untersuchende "Statistiker jemand ist, dessen Praxis durch die Theorie bestimmt ist. Ohne Theorie

---

1) Vgl. u.a. Anderson, Oskar (1949): S.10-14 u. S.20-23; Deming, Edwards W. (1960): S.26; Désabie, J. (1966): S.5ff.; Sampford, M.R. (1962): S.1-4; Vance, Lawrence L./Neter, John (1956): S.12/13; Zarkovich, Slobodan S. (1961): S.1-23.

2) Burnham, James (1949): S.178.

3) Z.B. Krebsforschung im Sinne der Zigarettenindustrie, s. Randow, Thomas v. (1964) in: DIE ZEIT v. 13.März 1964.

ist keine rationale Wahl der Methode möglich."[1] Aus diesem Grund muß auf einen detaillierten Stichprobenplan größten Wert gelegt werden. Die Elemente eines Stichprobenplans[2], auf die wir in den folgenden Abschnitten näher eingehen wollen, sind:

1. Die Stichprobenbasis. Hier muß die Beschaffenheit des vorliegenden Materials geprüft werden: die Grundgesamtheit muß genau abgegrenzt werden; es wird gefragt, welche Informationen sind vorhanden und welche Information wollen wir erhalten.
2. Das Stichprobenmodell, das sich aus dem Auswahlmodell und der Hoch- und Fehlerrechnung zusammensetzt. Es liefert uns die Schätzwerte mit den Vertrauensgrenzen, die das Ziel der Untersuchung sind.
3. Die Stichprobendurchführung
   a) Das Optimalprinzip oder das wirtschaftliche Prinzip. Die gewünschte Information soll mit den geringsten Kosten, bzw. bei gegebenen Kosten soll die größtmögliche Information geliefert werden. Meist läuft es darauf hinaus, ein Minimum der Varianz bei gegebenen Kosten zu erhalten[3].
   b) Kontrolle über Nichtstichprobenfehler, deren Ursachen aufgedeckt und deren Einfluß beseitigt werden sollen.
   c) Zur gewissenhaften Vorbereitung einer Stichprobe gehört eine Probeerhebung, die eine Reihe von Fehlerquellen ausmerzen hilft und frühzeitig Anhaltspunkte über die Wirksamkeit der Stichprobe liefert.

2.2.2 Die Stichprobenbasis

Die Basis der Stichprobe umfaßt das gesamte physische Material,

---

[1] Deming, Edwards W. (1960): S.25, eigene Übersetzung.

[2] Vgl. hierzu Deming, Williams E. (1960): S.38/39 und Dalenius, Tore (1962): S.21ff..

[3] Vgl. Hájek, Jaroslav (1958): S.130ff..

das uns für eine Untersuchung vorliegt. Es enthält: Adressenmaterial, Karteien, Gebietskarten, schon vorhandene Statistiken, aber auch institutionelle Begrenzungen (z.B. Statistisches Gesetz), nicht zu überschreitende Kostenvoranschläge, ferner Möglichkeiten der Erhebung, wie direkte Beobachtung, Anschreiben, Interviewen, Telefonumfrage in der Großstadt, vorhandene ausgebildete Moniteure (wie in Belgien) und zuletzt die Möglichkeiten der Auswertung (moderne Rechenanlagen), kurz alle informativen, wirtschaftlichen und technischen Möglichkeiten. Dazu müssen die Grundgesamtheit wie die Stichprobenelemente genau abgegrenzt werden. Aus der Basis ergibt sich dann die anwendbare Stichprobenmethode.[1)]

So empfiehlt sich eine mehrstufige Stichprobe, wenn das Adressenmaterial gering ist. Aus dem gleichen Grund empfiehlt sich bei Haushaltsstichproben die Methode von Kish, bei der der Interviewer aus dem ausgewählten Haushalt einen Erwachsenen befragt, den er per Zufall trifft. Bei einer vollständigen Kartei bereitet die Auswahl keine Schwierigkeiten, dagegen müssen bei unvollständiger und veralteter Basis entsprechende Methoden gefunden werden.[2)] Ferner ist aus der vorliegenden Basis zu ersehen, ob eine Schätzung mit systematischer Auswahl oder einer Auswahl mit ungleichen Wahrscheinlichkeiten angebracht ist. Sie zeigt auf, ob zusätzliche Informationen vorliegen, die wirksamere Schätzmethoden, wie A-posteriori-Schichtung, Verhältnis-, Regressions- oder Differenzenschätzung zulassen. Der erste Schritt jeder Stichprobenerhebung ist also die Analyse der Basis der Stichprobe.

Einige immer wieder auftretende Miβverständnisse, die Basis der Stichprobe betreffend, sollen im folgenden angedeutet werden.[3)]

- In einer Zufallsstichprobe dürfen keine typischen Fälle zur "Verbesserung" angeführt werden.
- Nicht die Urteilskraft, sondern die Theorie ist die

---

1) Vgl. Désabie, J. (1966): S.103ff..

2) Vgl. Désabie, J. (1966): S.114ff., ebenso: Zarkovich, S.S. (1964): S.116ff..

3) S. Deming, Edwards W. (1960): S.28ff..

Grundlage einer Stichprobenerhebung.

- Nicht die Größe der Stichprobe ist ein Kriterium für die Genauigkeit, sondern die Auswahl- und Schätzmethode, sowie die Wahl der Erhebungseinheit.
- Der Zufall kann nicht durch einen größeren Stichprobenumfang ersetzt werden.
- Nicht die absolute Zahl der Stichprobeneinheiten sagt etwas über die Genauigkeit aus, sondern der Anteilssatz.
- Die Qualität einer Stichprobe ist kein Glückstreffer, sondern sie ist eingebaut.
- Eine Stichprobe ist nicht die letzte Zuflucht, wenn eine Vollzählung nicht möglich ist, sondern die erste.
- Die Unsicherheit ist ein objektives Maß in der modernen Stichprobenmethode.
- Man kann Meinungen und Quantitäten mit Stichproben messen.
- Übereinstimmung (gute oder schlechte) mit Vollzählungen sagt nichts über die Genauigkeit der Stichprobe oder über die Qualität des Zensus.

Aus dem Gesagten ergibt sich, daß bei Schätzungen nur rationale Komponenten eine Rolle zu spielen haben, und eine solche Komponente des Stichprobenplans ist das Stichprobenmodell.

2.2.3 Das Stichprobenmodell

Von größter Bedeutung für den Theoretiker ist das Stichprobenmodell. Folgende Systematik, die im Institut für Wirtschafts- und Sozialstatistik, Tübingen, entworfen wurde, soll die Zusammenhänge deutlich machen (Abb. 4).

Verschiedene oder auch nur ein Auswahlverfahren bilden das Auswahlmodell. Es kann hier mit variabler oder nicht variabler Auswahlwahrscheinlichkeit gearbeitet werden; ferner können einfache, geschichtete, mehrstufige, mehrphasige oder/und kombinierte Auswahlverfahren in Anwendung gebracht werden. Zusammen mit der Hoch- und Fehlerrechnung ergibt es so das Stichprobenmodell.

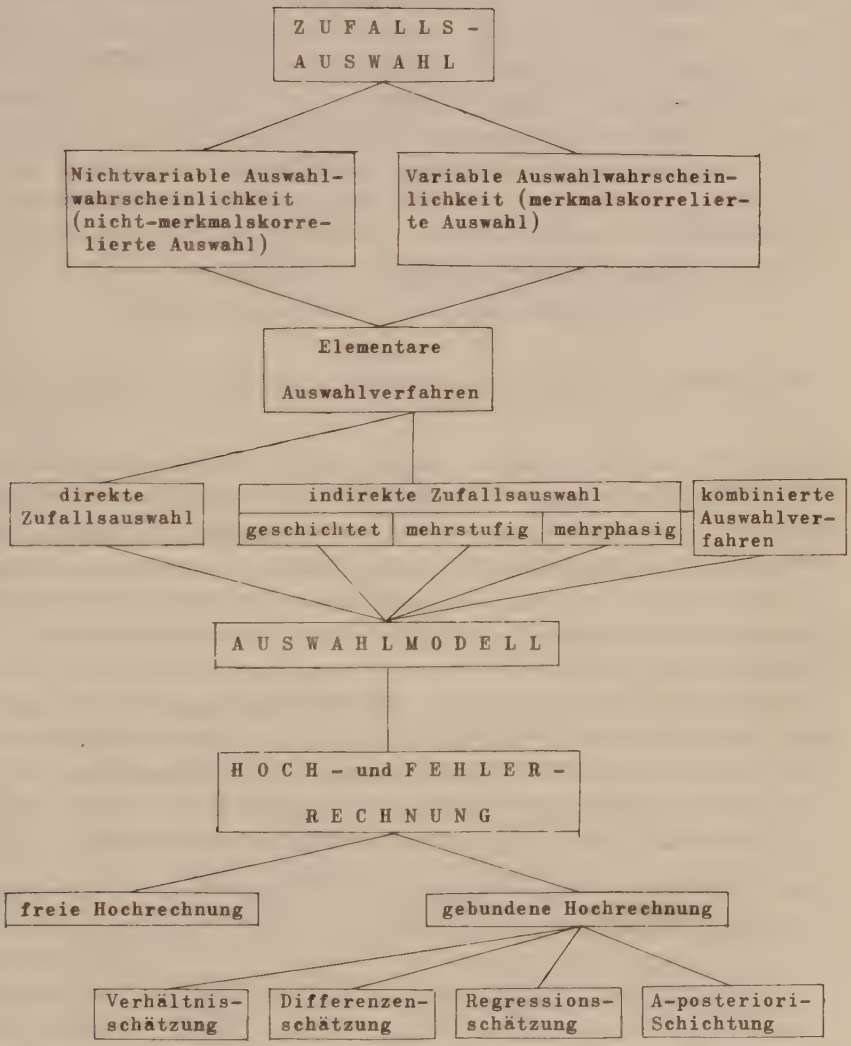


Abb.4 Schematische Darstellung des Stichprobenmodells

Quelle: Institut für Wirtschafts- und Sozialstatistik, Universität Tübingen.

Die verschiedenen Modelle wurden teils auf theoretischer Grundlage erarbeitet, wie z.B. die einfache Zufallsauswahl, teils von der Praxis beeinflußt, wie z.B. die Klumpenauswahl und die mehrstufige Auswahl, wo der Mangel an Adressenmaterial das Verfahren nahelegte.[1] Zumeist handelte es sich um ein fruchtbares gegenseitiges Beeinflussen von Theorie und Praxis, woraus dann neue Verfahren für die Stichprobenerhebungen entstanden und weiterhin entstehen werden. Hier sollen die einzelnen Modelle nicht näher untersucht werden, außer natürlich das spezielle Hochrechnungsmodell der A-posteriori-Schichtung.

2.2.4 Die Stichprobendurchführung

Bei der Durchführung einer Stichprobe gelten folgende Eigenschaften für ein gutes Programm[2]: Die Erhebung muß nützlich und von umfassendem Inhalt sein; sie muß zuverlässige Ergebnisse liefern; die Klassifikationen und Definitionen müssen verständlich und eindeutig sein; geringer Zeitaufwand und Wirtschaftlichkeit müssen garantiert sein, ebenso wie am Schluß die genaue Interpretation und Darstellung. Um diese Eigenschaften gewährleisten zu können, ist eine Probeerhebung äußerst wichtig. Probeerhebungen sind geboten bei neuen Erhebungsgegenständen und/oder neuen Erhebungsmethoden.[3] Sie zeigen Fehlerursachen auf, bringen neue Erkenntnisse und geben die Möglichkeit, die Wirksamkeit der Erhebung im groben im voraus abzuschätzen.

Jede Erhebung, ob Totalzählung oder Stichprobe, ist mit Fehlern behaftet. Der Fehler wird üblicherweise definiert als Differenz zwischen dem wahren und dem geschätzten Wert. Der Gesamtfehler einer Stichprobenerhebung setzt sich zusammen aus dem Stichprobenfehler und dem systematischen Fehler. Dieser Gesamtfehler läßt sich in der bekannten Form darstellen:[4]

---

1) Sukhatme, P.V. (1966): S.374ff..

2) Deming, William Edwards (1966): S.3/4.

3) Vgl. Statistisches Bundesamt (Hrsg.)(1960): S.104/105. Hier werden auch eine Reihe von Beispielen für Probeerhebungen angegeben, so S.145ff., S.258ff., usw..

4) S. Kish, Leslie (1965): S.514.

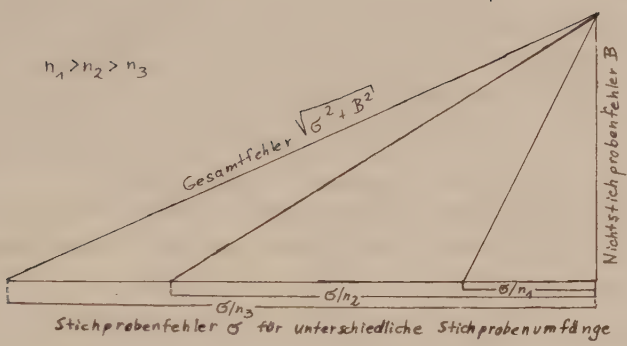


Abb.5 Darstellung des Gesamtfehlers bei Stichprobenerhebungen

Der Stichprobenfehler läßt sich mit größerem Stichprobenumfang und/oder verbesserten Schätzverfahren verringern. Dagegen sind Nichtstichprobenfehler unabhängig von n. Sie werden am besten durch Probeerhebungen und Nachuntersuchungen gefunden. Die bekanntesten Fehlerarten hier sind:[1)]

- Stichprobenmethoden, die einen Bias aufweisen (Biased procedure)
- Vergessen von Merkmalen
- Unverständlichkeit und Nichteindeutigkeit der Definitionen und Kategorien; z.B. darf bei Viehzählungen nicht nur stehen: Ferkel und Jungschweine, sondern es müssen genaue Altersangaben, wie unter 2 Monate alt, 2 bis unter 4 Monate alt, usw. stehen, da sonst leicht Spaltenverschiebungen entstehen.[2)]
- Beeinflussung durch den Interviewer muß in jeder Hinsicht ausgeschlossen werden[3)]. Der Interviewer darf keine eigenen, noch so gut gemeinten Initiativen er-

---

1) Vgl. Deming, William Edwards (1966): S.26-52; Hendricks, Walter A. (1956): S.322ff.; Sukhatme, P.V. (1966): S.386ff. und Zarkovich, S.S. (1964): S.5ff..

2) Strecker, H./Steylaerts, R. (1966): S.78.

3) Vgl. Noelle, Elisabeth (1963): S.44ff..

greifen. So wurden z.B. bei den Nachprüfungen der Viehzählung in Belgien selbst von geschulten Moniteuren bei Betrieben, die aus irgendeinem Grund keinen Viehbestand mehr hatten, keine Ersatzbetriebe genommen. Da bei einem solchen Nullbetrieb kein Fehler gemacht werden kann, kann auch keiner nachgeprüft werden, und der Stichprobenumfang wird automatisch reduziert.[1)]

- Non-response-Probleme[2)]
- Alle übrigen Nichtstichprobenfehler, wie Rechenfehler und ähnliches.

Jede Erhebung allgemein, wie jede Reduzierung von Fehlern im besonderen ist mit Kosten verbunden. Daraus ergibt sich die Frage der Wirtschaftlichkeit, das sogenannte Optimierungsproblem. Kosten entstehen in allen drei Phasen einer statistischen Erhebung.

Bei der Planung und Vorbereitung handelt es sich meist um fixe Kosten (C_o). Bei der Erhebung entstehen variable Kosten (C(n)), die abhängig sind vom Stichprobenumfang n; es entstehen Reise-, Porto-, Interviewerkosten, usw.. Bei der Aufbereitung handelt es sich zum überwiegenden Teil um fixe Kosten, aber auch - bei Rechenkosten, der Größe von Tabellen usw. - um variable Kosten. Die allgemeine und einfachste Form der Kostenfunktion sieht folgendermaßen aus

$$C = C_o + C(n)$$

Würden wir den Genauigkeitsgewinn G(n) in Geldeinheiten ausdrücken, erhielten wir das Maximum einer Informationsaussage,[3)] wenn der zusätzliche Genauigkeitsgewinn in Geldeinheiten gleich den zusätzlichen Kosten ist:

$$\Delta G(n) = \Delta C(n) \quad \text{bzw.} \quad G'(n) = C'(n)$$

---

1) Dies ergab sich bei der Arbeit über die "Feststellung von Angabefehlern bei Viehzählungen in Belgien" für die EWG am Institut für Wirtschafts- und Sozialstatistik in Tübingen.

2) S. Kish, Leslie (1965): S.532ff..

3) Vgl. hierzu Cochran, William G. (1964): S.82.

Dabei muß der Zeitaufwand als vorgegeben betrachtet werden.[1] Der vorgegebene Zeitaufwand ist dann als Nebenbedingung anzusehen. In der praktischen Erhebung interessiert aber meist die maximale Genauigkeit bei gegebenen Kosten (Optimum). Ferner wird nach den durchschnittlichen Kosten je Einheit gefragt, worin auch die fixen Kosten berücksichtigt sind, und nicht nach den marginalen Kosten.

Obwohl Kostenanalysen in theoretischer und praktischer Hinsicht gefordert werden,[2] findet man in den Lehrbüchern allenfalls kurze theoretische Analysen. Für einzelne Methoden werden optimale Werte angegeben[3] und gesagt, für kombinierte Methoden gelten die gleichen Bedingungen. Ein Abwägen verschiedener kombinierter Methoden vom Kostengesichtspunkt aus wird selten gemacht.[4] Das bedeutet für jeden, der eine Stichprobenerhebung durchführt, daß er eigene Kostenuntersuchungen anstellen muß. Er muß entscheiden,

- ob er den Stichprobenumfang oder die Stichprobeneinheit aus Kostengründen vergrößern soll (z.B. statt Einzelbefragungen nun eine Haushaltsbefragung durchführen);
- ob die Verringerung der Kosten bei Stufenverfahren nicht durch den Stufeneffekt wettgemacht wird;
- ob Hochrechnungsmethoden mit zusätzlicher Information (gebundene Hochrechnung) die Rechenkosten nicht zu stark erhöhen und ob ein eventueller Bias die Genauigkeit wieder vermindert usw..

Das Kostenproblem ist bis heute allgemein noch nicht gelöst und kann vielleicht auch nicht allgemein gelöst werden. Jedoch würden praktische Kostenanalysen Kostenwerte liefern, die bei der Planung einer Stichprobenerhebung als Richtzahlen eingesetzt werden könnten.

---

1) Vgl. Szameitat, Klaus/Schäffer, Karl-August (1964): S.129.

2) United Nations (Hrsg.)(1964): S.4.

3) Vgl. Sampford, M.R. (1962); Yates, Frank (1965): S.283ff..

4) Eine Ausnahme bildet hier Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956b): S.299ff., der ein Beispiel dieser Art rechnet; ebenso sind zwei Beispiele durchgeführt bei: Szameitat, Klaus/Schäffer, Karl-August (1964): S.140ff..

2.3 Die bedeutenden Hochrechnungsarten neben der A-posteriori-Schichtung

In diesem Abschnitt wollen wir den Rahmen aufzeigen, in dem die A-posteriori-Schichtung unter den Schätzverfahren steht. Neben dieser kurzen Beschreibung der wichtigsten Hochrechnungsarten im folgenden, werden diese später zu Vergleichen der Wirksamkeit mit der A-posteriori-Schichtung herangezogen, und es sollen mit ihnen kombinierte Verfahren mit der A-posteriori-Schichtung gebildet werden.

2.3.1 Allgemeine Einteilung

Man unterscheidet bei den Hochrechnungsmethoden zwei Gruppen: freie (einfache) Hochrechnung und gebundene Hochrechnung.[1)]

Die freie Hochrechnung ist nichts anderes als ein "Aufblähen" des errechneten Mittelwertes aus einer Stichprobe auf den Totalwert. Dieser ist für den Konsumenten der Statistik oft von größerem Interesse als der Mittelwert, z.B. interessiert ihn nicht nur der durchschnittliche Ernteertrag, sondern der Schätzwert der gesamten Ernte. Kennzeichnend für die freie Hochrechnung ist, daß man einfach die Zahl der Stichprobeneinheiten extrapoliert und daß "die Stichprobenbasis den zu extrapolierenden Mittelwert selber liefert"[2)].

Bei der gebundenen Hochrechnung werden zusätzliche Informationen zur Schätzung benutzt, z.B. die Daten einer Vollzählung, Stichprobenmittel oder -totalwerte, die ein eng korreliertes Merkmal betreffen (z.B. wird zur Schätzung der Apfelernte als zusätzliche Information der Apfelbaumbestand genommen); aber auch Ergebnisse aus früheren Erhebungen über den Untersuchungsgegenstand sind sehr gute Lieferanten zusätzlicher

---

1) So die Terminologie des Statistischen Bundesamtes, s. Statistisches Bundesamt (Hrsg.)(1960): S.46.

2) Thionet, M.P. (1953b): S.99.

3) Menges, Günter (1959): S.141.

Informationen. Diese Methoden werden in der angelsächsischen und französischen Literatur "als Schätzungen mit zusätzlichen Kenntnissen (estimation with supplementary information) bezeichnet, dies deswegen, weil bei diesen ... Verfahren in der Regel Kenntnisse verwendet werden, die aus anderen Quellen als der Stichprobe selbst stammen."[1] Die Schätzwerte werden auf einer verbreiterten Basis errechnet. Man beschränkt sich nicht mehr auf die einfache Stichprobenbasis; mit Hilfe zusätzlicher Informationen schaffen wir eine Art neue Basis der Extrapolation[2].

Auf dieser verbreiterten Basis werden dann bessere, aber offensichtlich auch kompliziertere Schätzformeln konstruiert. Die bekanntesten gebundenen Hochrechnungsarten neben der A-posteriori-Schichtung (Verhältnisschätzung, Regressionsschätzung und Differenzenschätzung) werden aus Gründen des späteren Vergleichs in den folgenden Abschnitten kurz beschrieben.

2.3.2 Freie oder einfache Hochrechnung

Die freie Hochrechnung basiert auf der einfachen Zufallsstichprobe. Ist der Auswahlsatz n/N, so ist der Hochrechnungsfaktor N/n. Somit erhalten wir für den Schätzwert des Totalwertes:

$$\hat{Y} = N \cdot \bar{y} = \frac{N}{n} \sum_{i=1}^{n} y_i$$

Bilden wir den Erwartungswert, so finden wir, daß $N \cdot \bar{y} = \frac{N}{n} \sum y_i$ ein unverzerrter Schätzwert für den gesamten Merkmalsbetrag der Grundgesamtheit ist.

Die Varianz des Totalwertes ist einfach das N^2-fache der Varianz des Stichprobenmittelwertes. Also

$$V(\hat{Y}) = \frac{N^2 S^2}{n} \frac{(N-n)}{N} = \frac{N(N-n)}{n} \cdot S^2 = \frac{N^2(N-n)}{N-1} \frac{\sigma^2}{n}$$

---

1) Menges, Günter (1959): S.141.

2) Thionet, M.P. (1953b): S.99.

Die Standardabweichung für den Totalwert ist dann

$$\sigma_{\hat{Y}} = \frac{N \cdot S}{\sqrt{n}} \sqrt{\frac{N-n}{N}}$$

Für den Schätzwert der Varianz des Totalwertes gilt das gleiche wie für den des Stichprobenmittelwertes. Der unverzerrte Schätzwert der Varianz des Totalwertes ist also:

$$v(\hat{Y}) = s^2_{\hat{Y}} = \frac{N^2 s^2}{n} \frac{N-n}{N}$$

Davon ist die Standardabweichung[1]:

$$s_{\hat{Y}} = \frac{N \cdot s}{\sqrt{n}} \sqrt{\frac{N-n}{N}}$$

Die Vertrauensgrenzen, die wir aus der Elementartheorie kennen, sind dann wie folgt:[2]

$$\hat{Y} - \frac{tNs}{\sqrt{n}} \sqrt{\frac{N-n}{N}} \leq Y \leq \hat{Y} + \frac{tNs}{\sqrt{n}} \sqrt{\frac{N-n}{N}}$$

2.3.3 Verhältnisschätzung

2.3.3.1 Methode

Man kann häufig beobachten, daß die Verhältnisse zweier Größen in einer Gesamtheit weniger variieren als die einzelnen Größen selber. In dieser Beobachtung liegt die Begründung für die Verhältnisschätzung, ebenso wie für die Quotientenmethode in der Zeitreihenanalyse. Darüber hinaus ist bei Schätzungen oft das

---

1) Wir wollen uns hier an die konventionellen Lehrbücher halten, die eine geringfügige Ungenauigkeit begehen, indem sie nicht darauf hinweisen, daß die Standardabweichung als Quadratwurzel aus der Varianz kein unverzerrter Schätzwert ist, sondern daß als Näherungswert bei genügend großem N (größer als 252) im Nenner statt N-1 der Wert N-1,5 stehen müßte. Vgl. Cureton, Edward E. (1968): S.22.

2) Statt des t-Faktors kann auch das international gebräuchliche Symbol $\xi_{1-\alpha/2}$ für die kritischen Werte der Normalverteilung verwendet werden.

Verhältnis zweier Charakteristika von großem Interesse[1].

Die Schätzformeln bei der Verhältnisschätzung sehen folgendermaßen aus, für den Mittelwert:

$$\bar{y}_v = \frac{\bar{y}}{\bar{x}} \bar{X} = f \cdot \bar{X}$$

für den Totalwert:

$$\hat{Y}_v = \frac{\bar{y}}{\bar{x}} X = f \cdot X$$

Wir bilden also das Verhältnis der zwei Variablen y und x, die - um sinnvolle Ergebnisse zu erhalten - sachlich korreliert sein müssen. Ferner nehmen wir zur Schätzung noch den Mittelwert $\bar{X}$ bzw. den Totalwert X der zusätzlichen Information, die uns die Variable x liefert.

Die Verhältnisschätzung ist eine verzerrte Schätzmethode (biased estimate), und nur unter bestimmten Bedingungen liefert sie erwartungstreue Schätzwerte. Die Verzerrung rührt von dem Verhältnis f = y/x her; denn es gilt:

$$y = f \cdot x$$

Der Erwartungswert davon ist:

$$Ey = EfEx + Cov(f,x)$$

Ist der Korrelationskoeffizient gleich 0, entfällt die Kovarianz, und es ist:

$$Ey = EfEx = \frac{Y}{X} X = Y \qquad \text{(wahrer Wert)}$$

Je nach der Art des Korrelationskoeffizienten - ϱ kann negativ oder positiv sein - ist der Schätzwert kleiner oder größer als der wahre Wert.

In der Praxis ist der Bias meist eine vernachlässigbare Größe, wenn zur Schätzung genügend viele Stichprobeneinheiten genom-

---

[1]) Vgl. Murthy, M.N. (1963): S.746.

men werden (Cochran[1]) gibt hier für den Stichprobenumfang n>30 an). Tatsächlich ist f nicht verzerrt, wenn die "Gerade der besten Anpassung" durch den Ursprung geht.[2] Es ist nicht notwendig, genau auszuführen, was diese Gerade bedeutet, meist genügt es, eine Gerade nach Augenmaß durch eine Punkteschar eines Koordinatensystems zu legen, so daß sie annähernd durch die Mitte dieser Punkteschar geht. Wenn die Gerade durch den Ursprung geht, liegt kein Bias vor. Das läßt sich leicht aus der Form der Schätzfunktion herleiten. Die Form der Schätzfunktion $y = ax$ deutet geometrisch ausgedrückt auf eine Ursprungsgerade hin. Fällt die "Gerade der besten Anpassung" nun nicht mit der Geraden der Schätzfunktion der Verhältnisschätzung zusammen, d.h., zwingen wir die "Gerade der besten Anpassung", die dann keine der besten Anpassung mehr ist, durch den Ursprung, erhalten wir in den Schätzwerten natürlich eine Verzerrung.

Es ist leicht einzusehen, daß je größer die Punkteschar im Koordinatensystem ist, um so eher die "Gerade der besten Anpassung" durch den Ursprung geht. Hier ist es nicht so wie bei der Regressionsschätzung, daß der falsche Ansatz - nämlich die Form der Regressionsfunktion - unverändert bleibt. Mit zunehmendem Stichprobenumfang wird der Ansatz sozusagen verbessert. Cochran[3] hat an einem Beispiel die Wirksamkeit der Verhältnisschätzung demonstriert, indem er die Einwohnerzahl von Großstädten mit Hilfe der Verhältnisschätzung und der einfachen Hochrechnung geschätzt hat. Das nachstehende Schaubild zeigt das Ergebnis aus 200 Zufallsstichproben, oben die Verhältnisschätzung, darunter die einfache Hochrechnung; dabei bezeichnet x den wahren Wert.

---

1) Cochran, William G. (1964): S.157.
2) Vgl. hierzu Deming, W. Edwards (1960): S.425/426.
3) Cochran, William G. (1964): S.155ff..

Abb.6 Experimenteller Vergleich von Verhältnisschätzung und einfacher Hochrechnung

Quelle: Cochran, William G. (1964): S.156.

2.3.3.2 Formeln für Varianz und Bias

Die wesentlichen Schwierigkeiten der Theorie rühren von der Zufallsgröße x her, die im Nenner steht, da bei der einfachen Hochrechnung aus der Variablen y keine Schwierigkeiten entstehen. Die Verteilung[1)] der Verhältnisschätzung ist schwer zu behandeln, weil beide Variablen y und x des Verhältnisses f von Stichprobe zu Stichprobe variieren. Für den Bias und die Stichprobenvarianz der Schätzung besitzen wir keine exakten Formeln, sondern ausschließlich Approximationen. Daher gelten die folgenden Ergebnisse nur unter der Bedingung, daß der Stichprobenumfang groß genug ist, so daß a) das Verhältnis f annähernd normalverteilt ist und b) die "large-sample"-Formel für die Varianz gilt.

Mängel in der Theorie sind das Fehlen a) einer gut begründeten Formel, um die Frage zu beantworten: Wann ist der Stichproben-

---

1) Vgl. Cochran, William G. (1964): S.157.

umfang groß genug?, und b) einer handlichen Methode, Vertrauensgrenzen für kleine Stichproben zu schätzen. Als empirische Regel, ab wann "large-sample"-Ergebnisse angewandt werden können, wird bei Cochran eine Stichprobengröße von mindestens 30 Einheiten angegeben; ferner müsse die Stichprobe groß genug sein, damit beide Variationskoeffizienten von $\bar{x}$ und von $\bar{y}$ kleiner als 10% seien. Diese Regel sei bis jetzt jedoch noch ziemlich ärmlich begründet.

Der Bias ist definiert als die Differenz zwischen der mathematischen Abweichung des Schätzwertes und dem wahren Wert, also:

$$B(\bar{y}_v) = E(\bar{y}_v) - \bar{Y} = E(\frac{\bar{Y}}{\bar{x}}\bar{X} - \bar{Y})$$

Durch Umformungen und mit Hilfe der Entwicklung in eine abgebrochene Taylor-Reihe erhalten wir:

$$B(\bar{y}_v) = E(1 - \frac{\bar{x}-\bar{X}}{\bar{X}})(\bar{y} - \frac{\bar{Y}}{\bar{X}}\bar{x})$$

Daraus erhalten wir die Formel für den Bias

$$B(\bar{y}_v) = \frac{N-n}{N \cdot n}\frac{1}{\bar{X}}(FS_x^2 - \varrho S_x S_y)$$

dabei ist

$$F = \frac{\bar{Y}}{\bar{X}} .$$

Im Kalkül der Varianz wird der Bias systematisch vernachlässigt, man setzt näherungsweise $E(\bar{y}_v) = \bar{Y}$. Die Berechtigung dafür wird von der Größe des Bias hergeleitet, der bei der Varianz von der Ordnung $1/n^2$ ist. Zur Berechnung der Varianz führen wir die Hilfsvariable $u_i = y_i - F \cdot x_i$ ein und wenden für diese das Theorem für die Varianz des Mittelwertes einer einfachen Zufallsstichprobe an. Wir erhalten dann

$$V(\bar{y}_v) = V(\bar{u}) = \frac{N-n}{N \cdot n}\frac{\sum_1^N (u_i - \bar{U})^2}{N-1}$$

Dabei ist $\bar{U}$ der Mittelwert der Grundgesamtheit der Variablen u_i

und somit $\bar{U} = \bar{Y} - F \cdot \bar{X}$. Nach der Definition von $F = \frac{\bar{Y}}{\bar{X}}$ aber wird $\bar{U} = 0$, folglich

$$V(\bar{y}_v) = \frac{N-n}{N \cdot n} \frac{\sum^{N}(y_i - Fx_i)^2}{N-1}$$

Durch Umformungen erhalten wir dann die bekannte Formel der Varianz für die Verhältnisschätzung:

$$V(\bar{y}_v) = \frac{N-n}{N \cdot n} (S_y^2 + F^2 S_x^2 - 2FS_{xy})$$

bzw. mit Einführung des Korrelationskoeffizienten ϱ erhalten wir:

$$V(\bar{y}_v) = \frac{N-n}{N \cdot n} (S_y^2 + F^2 S_x^2 - 2F\varrho S_x S_y)$$

Um zum unverzerrten Schätzwert der Varianz der Verhältnisschätzung zu gelangen, suchen wir die unverzerrten Schätzwerte der einzelnen Komponenten der Varianz der Grundgesamtheit.[1] Wir erhalten dann für den Schätzwert der Varianz:

$$v(\bar{y}_v) = \frac{N-n}{N \cdot n} (f^2 s_x^2 + s_y^2 - 2fs_{xy})$$

2.3.3.3 Bias und Standardabweichung

An Hand dieses Verhältnisses: Bias zu Standardabweichung $(B_{(\bar{y}_v)} / \sqrt{V(\bar{y}_v)})$ kann man sozusagen a priori ersehen, ob ein auftretender Bias gegenüber der Streuung für die Hochrechnung vernachlässigbar ist, d.h., ob man den Bias mit Hilfe der Varianz in den Griff bekommt. Wir bilden also dazu das Verhältnis

$$\frac{B(\bar{y}_v)}{\sqrt{\frac{N-n}{N \cdot n} \sqrt{S_y^2 + F^2 S_x^2 - 2F\varrho S_x S_y}}} = \frac{\frac{N-n}{N \cdot n} \frac{1}{\bar{X}} (FS_x^2 - \varrho S_x S_y)}{\sqrt{\frac{N-n}{N \cdot n} \sqrt{S_y^2 + F^2 S_x^2 - 2F\varrho S_x S_y}}}$$

---

[1] Vgl. Deming, W. Edwards (1960): S.438/439.

Durch einige Umformungen erhalten wir (dabei nehmen wir den Betrag des Bias absolut):

$$\frac{\left|B(\bar{y}_v)\right|}{\sqrt{V(\bar{y}_v)}} \leq \frac{\sigma_{\bar{x}}}{\bar{X}}$$

Cochran[1] hat nun gezeigt: Ist der Variationskoeffizient ($\sigma_{\bar{x}}/\bar{X}$) - also einer Größe, die man bei der Hochrechnung aus der zusätzlichen Information gewinnt und vor der Berechnung der Stichprobe kennt - kleiner als 0,2, so kann man den Bias im Vergleich zur Größe der Standardabweichung ohne nennenswerte Genauigkeitseinbußen vernachlässigen. Voraussetzung ist, daß n groß genug ist, damit die Mittelwerte normalverteilt sind. Wie wir schon gesehen haben, nimmt der Bias bei zunehmender Stichprobengröße ab; bei großem n braucht der Bias ebenfalls nicht berücksichtigt zu werden. Um die Größe des Bias bei der Verhältnisschätzung festzustellen, gibt es verschiedene Wege. Eine nützliche Annäherung an das Problem sind Untersuchungen über die empirische Größe des Bias in Erhebungen der Vergangenheit. So haben Kish und Namboodiri bei einer analytischen Studie am "Survey Research Centre of the University of Michigan" gefunden, daß das Verhältnis von Bias zu Standardabweichung durchschnittlich 1% beträgt und selten größer als 4% wird.[2]

Entgegen dieser Erfahrungstatsache bleibt die theoretische Möglichkeit bestehen, daß der Bias der Verhältnisschätzung aus einer geschichteten Grundgesamtheit von einer Größe ist, die nicht vernachlässigt werden kann. Das Interesse an einer Eliminierung des Bias oder der Entwicklung einer unverzerrten Verhältnisschätzung ist daher verständlich. Ein Vorschlag stammt von Quenouille[3] (1949), der die Methode von Unterstichproben (subsample) benutzt. Dabei soll f das Ergebnis einer Stichprobe vom Umfang n sein und f_1 und f_2 das Ergebnis zweier Stichproben mit n/2 Beobachtungen, die man durch Aufspaltung der ursprüng-

---

1) Cochran, William G. (1964): S.14.
2) Vgl. Zarkovich, S.S. (1964): S.86.
3) Durbin, J. (1959): S.477-480.

lichen Stichprobe in zwei gleiche Teile erhält. Das Verhältnis wird durch $f_o = 2f - \frac{1}{2}(f_1 + f_2)$ gegeben; der Bias ist dabei von der Ordnung n^{-2}.

Eine andere Methode stammt von Lahiri[1] (1951) und besteht in der Definition und Annahme einer solchen Auswahlmethode, durch die der Bias in der Verhältnisschätzung eliminiert wird.

Eine weitere Möglichkeit wurde von Hartley und Ross[2] (1954) aufgezeigt, indem der Bias in der Hochrechnungsformel mitberücksichtigt wird. Dieser Ansatz erscheint interessant, wir wollen deshalb im folgenden Abschnitt kurz darauf eingehen.

Ferner haben eine Reihe von Autoren[3] gezeigt, daß bei einfacher Zufallsauswahl aus einer unendlichen Grundgesamtheit die Verhältnisschätzung eine "beste lineare unverzerrte Schätzung"[4] ist, wenn zwei Bedingungen erfüllt sind:

a) die Regressionslinie, die das Verhältnis zwischen x und y darstellt, durch den Ursprung geht,
b) die Varianz von y um die Regressionslinie proportional zu x ist.

Unter diesen Annahmen erscheint es aber fraglich, ob diese Ausführungen zu einer optimalen Verhältnisschätzung für die praktische Stichprobenschätzung von großem Wert sind.

2.3.3.4 Eine Variante der Verhältnisschätzung

Neben der Schätzfunktion $\bar{y}_v = f \cdot \bar{X}$

mit $$f = \frac{\bar{y}}{\bar{x}}$$

können wir eine Schätzfunktion der folgenden Form bilden

---

1) Lahiri, D.B. (1951): S.133-140.

2) Hartley, H.O./Ross, A. (1954): S.270/271.

3) S. Cochran, William G. (1964): S.166/167, ebenso: Sukhatme, Pandurang V. (1963): S.154ff., ferner: Zarkovich, S.S. (1964): S.85/86.

4) Cochran, William G. (1964): S.166, eigene Übersetzung.

$$\bar{y}_v' = \bar{f}' \cdot \bar{X}$$

mit $$\bar{f}' = \frac{1}{n} \sum f_i'$$

und $$f_i' = \frac{y_i}{x_i}$$

Aus dieser Schätzformel $\bar{y}_v' = \bar{f}' \cdot \bar{X}$ haben Hartley und Ross einen unverzerrten Schätzwert abgeleitet.[1] Wir bilden den Erwartungswert:

$$E(\bar{y}_v') = E(\bar{f}') \cdot \bar{X}$$

$$= \frac{E(\bar{y})}{E(\bar{x})} \cdot \bar{X} - Cov(f', x)$$

$$= \bar{Y} - Cov(f', x)$$

Der Erwartungswert ist also kleiner, und zwar um die Kovarianz von (f', x) als der wahre Wert. Hartley und Ross nehmen daher zur Schätzung die Formel:

$$\bar{y}_v' = \bar{f}' \cdot \bar{X} + Cov(f', x)$$

Für den Kovarianzausdruck suchen wir den unverzerrten Schätzwert. Daraus ergibt sich als endgültige Schätzformel

$$\bar{y}_v' = \bar{f}' \cdot \bar{X} + \frac{n(N-1)}{N(n-1)} (\bar{y} - \bar{f}' \cdot \bar{x})$$

Obwohl diese Variante der Verhältnisschätzung unverzerrte Schätzwerte liefert, wird sie in der Literatur nur spärlich behandelt. Ein Grund dafür könnte sein, daß das Mehr an Rechenarbeit zur Berechnung jedes einzelnen Verhältnisses y_i/x_i durch die Verbesserung der Ergebnisse nicht ausgeglichen wird, da der Bias doch meist sehr gering ist.

---

[1]) Hartley, H.O./Ross, A. (1954): S.270/271.

2.3.3.5 Stichprobenauswahl mit der Wahrscheinlichkeit proportional zur Hilfsvariablen x

Die Methode ist folgende: Die zu untersuchende Variable y wird gemäß dem Gewicht ausgewählt, das dem Gewicht der mit ihr eng korrelierten Variablen x der Grundgesamtheit entspricht. Die Auswahlwahrscheinlichkeit p_i der y_i, die proportional zu x sein soll, läßt sich leicht errechnen. Es ist $p_i = c \cdot x_i$, wobei c der Proportionalitätsfaktor ist. Wegen der Bedingung

$$\sum_{i=1}^{N} p_i = 1$$

erhalten wir $$\sum_{i=1}^{N} p_i = \Sigma c x_i = c \Sigma x_i = 1$$

daraus $$c = \frac{1}{\sum_{i=1}^{N} x_i} = \frac{1}{N \cdot \bar{X}} .$$

Also ist die Auswahlwahrscheinlichkeit

$$p_i = x_i \frac{1}{N \cdot \bar{X}} .$$

Um die Schätzformel zu gewinnen, rechnen wir das Merkmal mit dem reziproken Auswahlsatz hoch und erhalten

$$\hat{Y}_i = \frac{y_i}{p_i} .$$

$\hat{Y}_i$ ist ein möglicher Schätzwert für den Totalwert Y. Einen Schätzwert für den Mittelwert $\bar{Y}$ erhalten wir dann, indem wir durch N teilen, also:

$$\bar{Y}_i = \frac{y_i / p_i}{N}$$

Von diesen möglichen Schätzwerten haben wir nun n Stichprobeneinheiten zur Verfügung. Diese wollen wir ausnützen und erhalten dann den Schätzwert der Verhältnisschätzung mit proportionaler Auswahl zu x:

$$\bar{y}_v' = \frac{1}{n} \sum \bar{Y}_i = \frac{1}{n} \sum \frac{y_i / p_i}{N}$$

Für p_i setzen wir den errechneten Wert ein und erhalten die endgültige Schätzformel:

$$\bar{y}_v' = \frac{1}{n} \sum \frac{y_i N \bar{X}}{x_i N} = \frac{1}{n} \sum \frac{y_i}{x_i} \bar{X}$$

$$= \bar{f}' \cdot \bar{X}$$

Diese Schätzfunktion liefert, im Gegensatz zu der im vorigen Abschnitt behandelten, unverzerrte Schätzwerte für $\bar{Y}$. Denn der Erwartungswert von $\bar{f}'$ weist keinen Bias auf, da nämlich "die Einheiten mit der Wahrscheinlichkeit proportional zu x ausgewählt worden sind, trägt jede Einheit der Gründgesamtheit zum Erwartungswert von $\bar{f}'$ im Verhältnis zu ihrer Auswahlwahrscheinlichkeit bei".[1] Die Erwartungstreue läßt sich leicht zeigen:

$$E(\bar{y}_v') = E(\bar{f}'\bar{X}) = \bar{X} E \bar{f}' = \bar{X} E \frac{y_i}{x_i} = \bar{X} \sum^N \frac{y_i}{x_i} p_i$$

$$= \bar{X} \sum^N \frac{y_i x_i}{x_i N \bar{X}} = \frac{1}{N} \sum^N y_i = \bar{Y}$$

2.3.3.6 Separate und kombinierte Verhältnisschätzung

Die Verhältnisschätzung soll nun mit dem geschichteten Auswahlverfahren kombiniert werden. Wir haben dann zwei Möglichkeiten:
1. Die <u>separate</u> Verhältnisschätzung. Dabei wird für jede einzelne Schicht die Verhältnisschätzung angewandt (separat), und diese Schichtgesamtheiten ($\hat{x}_h$) werden dann zum endgültigen Schätzwert $\bar{X} = \sum_h \hat{x}_h$ addiert. Die Schätzformel für den Totalwert ergibt dann

$$\hat{Y}_{vs} = \sum_h^L \frac{y_h}{x_h} X_h$$

(L = Anzahl der Schichten; x_h, y_h = Werte der h-ten Schicht)

---

[1]) Sampford, M.R. (1962): S.123, eigene Übersetzung.

2. Die kombinierte Verhältnisschätzung. Hier werden alle Schichten zusammen hochgerechnet (kombiniert). Aus den Stichprobendaten errechnet man in der üblichen Weise für die geschichteten Stichproben die Schätzwerte für X und Y. Die Formel für den Totalwert der kombinierten Verhältnisschätzung heißt also:

$$\hat{Y}_{vk} = \frac{\sum_h^L N_h \bar{y}_h}{\sum_h^L N_h \bar{x}_h} X$$

Ein Vergleich der beiden Methoden zeigt, daß die separate Verhältnisschätzung bessere Ergebnisse liefert, jedoch gilt das nur, solange der Umfang der Stichprobe in den einzelnen Schichten groß genug ist, damit die Anwendung der Approximationsformeln für die Varianz möglich ist.

2.3.4 Regressionsschätzung

2.3.4.1 Methode

Von den hier behandelten Hochrechnungsmethoden nützt die Regressionsschätzung die vorliegende Information am besten aus;[1] denn sie gibt die Möglichkeit, das theoretische Modell den gegebenen Merkmalen anzupassen. Man sucht bei der Regressionsschätzung diejenige Funktion, - z.B. geometrisch gesehen eine Gerade oder Parabel - die der Merkmalsverteilung am besten entspricht. Zwei Schaubilder sollen das verdeutlichen.

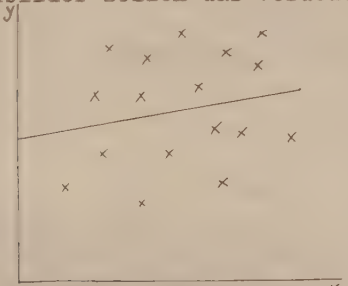


Abb.7 Gerade als Kurve der besten Anpassung

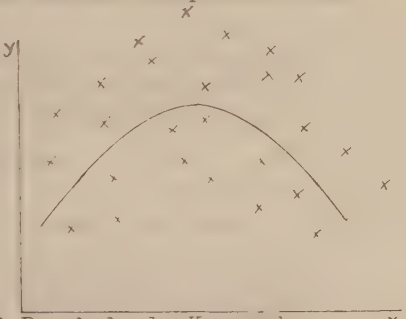


Abb.8 Parabel als Kurve der besten Anpassung

[1]) Vgl. Statistisches Bundesamt (Hrsg.)(1960): S.48.

Es liegt also nicht wie bei der Verhältnisschätzung von vornherein eine Gerade als Schätzfunktion fest. Deshalb scheint es mir fraglich, ob man sagen kann, daß "die Regressionsschätzung als Verallgemeinerung der Verhältnisschätzung angesehen werden kann"[1]. Die Methode ist in ihrer Grundkonzeption eine völlig andere, wenn auch der Form nach die Verhältnisschätzung wie eine "verkümmerte" Regressionsschätzung aussieht.[2] Ob zur Schätzung eine Gerade oder Parabel als Kurve der besten Anpassung genommen wird, ist von der theoretischen Behandlung her von untergeordneter Bedeutung; deshalb wollen wir uns in diesem kurzen Abriß auf die Gerade als Kurve der besten Anpassung beschränken.

Nachdem man die Verteilung der Wertepaare (x_i, y_i) untersucht hat, legt man die Art der Kurve - ob Gerade, Parabel, e-Funktion - fest. Dann bestimmt man mit Hilfe der Methode der kleinsten Quadrate[3] die Parameter der Funktion. Diese werden so bestimmt, daß die Summe der Quadrate der Ordinatenabstände der Punkte p_i von dieser Kurve zum Minimum wird. Dazu folgendes Schaubild:

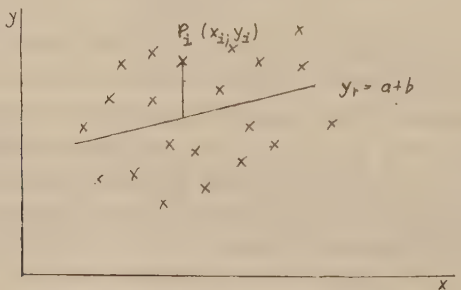


Abb.9 Verteilung der Wertepaare $P_i(x_i, y_i)$ um die Regressionsgerade

---

1) S. Hendricks, Walter A. (1956): S.298.

2) Diese Ansicht wurde z.B. in der Vorlesung über "Stichprobentheorie" im Wintersemester 1962/63 von Herrn Professor Dr. Heinrich Strecker vertreten.

3) Die Herleitung ist in jedem größeren Lehrbuch zu finden, z.B. Kellerer, Hans (1963): S.190ff..

Mit Hilfe von partiellen Ableitungen finden wir die Bestimmungsgleichungen für a und b, die sogenannten Normalgleichungen. Wir erhalten dann als Schätzfunktion zur Schätzung des Mittelwertes

$$\bar{y}_r = \bar{y} + b(\bar{X}-\bar{x})$$

bzw. zur Schätzung des Totalwertes

$$\hat{Y}_r = N \cdot \bar{y}_r = N\left[\bar{y} + b(\bar{X}-\bar{x})\right] .$$

Dabei erhalten wir wieder die zusätzliche Information aus der Variablen x.

2.3.4.2 Varianz der Regressionsschätzung

Die exakte Herleitung der Varianz ist für diesen kurzen Abriß zu umfangreich. Wir begnügen uns hier mit dem Näherungswert der Varianz. Dieser wird in der Literatur oft als Varianz der Regressionsschätzung angegeben.[1] Bei einem Stichprobenumfang von n>50 wird dieser Näherungswert für die Varianz als genügend angesehen.[2]

Wir nehmen folgende Strukturierung der Daten beim Regressionsmodell an[3]:

$$y_i = \bar{Y} + B(x_i-\bar{X}) + e_i$$

bzw. als Mittelwert in der Stichprobe

$$\bar{y} = \bar{Y} + B(\bar{x}-\bar{X}) + \bar{e}$$

B bedeutet den linearen Regressionskoeffizienten in einer endlichen Grundgesamtheit und ist definiert:

$$B = \frac{\sum_N (y_i-\bar{Y})(x_i-\bar{X})}{\sum_N (x_i-\bar{X})^2}$$

---

[1]) Unter anderen: Statistisches Bundesamt (Hrsg.)(1960): S.528; Dalenius, Tore (1957): S.20; Kapur, J.N./Saxena, H.C. (1963): S.221; Sampford, M.R. (1962): S.134.

[2]) S. Cochran, William G. (1964): S.198.

[3]) Cochran, William G. (1964): S.193.

e_i ist eine Restvariable mit $\sum_{i=1}^{N} e_i = 0$. "Dabei ist zu beachten, daß keine lineare Beziehung zwischen y_i und x_i vorausgesetzt wird. Die Grundgesamtheit besteht aus einer Reihe von N Wertepaaren (y_i, x_i), und die offenbar lineare Regression in der (obigen, S.H.) Gleichung wird herbeigeführt durch die Definitionen von B und e_i."[1] Zur Herleitung der Varianz setzen wir für den Schätzwert die Formel der Regressionsschätzung:

$$\bar{y}_r = \bar{y} + b(\bar{X}-\bar{x})$$

und für den wahren Wert die obige Gleichung ein, nämlich:

$$\bar{Y} = \bar{y} - B(\bar{x}-\bar{X}) - \bar{e}$$

Also haben wir:

$$\bar{y}_r - \bar{Y} = \bar{y} + b(\bar{X}-\bar{x}) - \bar{y} + B(\bar{x}-\bar{X}) + \bar{e}$$

$$= \bar{e} + (b-B)(\bar{X}-\bar{x})$$

Beim Näherungswert der Varianz vernachlässigen wir den Stichprobenirrtum, der dem Schätzwert b anhaftet; die Differenz b-B wird angenähert durch 0. Wir erhalten also die reduzierte Gleichung

$$\bar{y}_r - \bar{Y} = \bar{e}$$

Also wird der Näherungswert für die Varianz der Regressionsschätzung

$$E(\bar{y}_r - \bar{Y})^2 = E(\bar{e}^2) = \sigma^2_{\bar{e}}$$

Wir sehen sofort, daß das ein Varianzausdruck ist, da laut Definition $E(\bar{e})$ gleich 0 ist. Nun gilt bekanntermaßen bei endlicher Grundgesamtheit:

$$E(\bar{y}_r - \bar{Y})^2 = \sigma^2_{\bar{e}} = \frac{N-n}{N-1} \frac{\sigma^2_e}{n}$$

Nun suchen wir nicht die Varianz von e, sondern die Varianz in

---

[1]) Cochran, William G. (1959): S.142, eigene Übersetzung.

Ausdrücken der Variablen x und y. Laut Definition war

$$e_i = y_i - \bar{Y} - B(x_i - \bar{X})$$

Wenn wir diese Gleichung in die Varianz einsetzen, erhalten wir als Ausdruck für diese:

$$\sigma_e^2 = E\left[(y_i - \bar{Y}) - B(x_i - \bar{X})\right]^2$$

Mit Hilfe des Korrelationskoeffizienten können wir diese Gleichung umformen:

$$\varrho = \frac{E(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{E(x_i - \bar{X})^2 \; E(y_i - \bar{Y})^2}} = B\sqrt{\frac{E(x_i - \bar{X})^2}{E(y_i - \bar{Y})^2}}$$

oder

$$B = \varrho\sqrt{\frac{E(y_i - \bar{Y})^2}{E(x_i - \bar{X})^2}}$$

Wir erhalten dann:

$$\sigma_e^2 = E(y_i - \bar{Y})^2 - \varrho^2 E(y_i - \bar{Y})^2 = \sigma_y^2(1 - \varrho^2)$$

Setzen wir folgende Beziehung an:

$$\sigma_y^2 = \frac{N-1}{N} \cdot S_y^2$$

erhalten wir als Näherungswert für die Varianz des Mittelwertes der Regressionsschätzung:

$$V(\bar{y}_r) = \frac{N-n}{N \cdot n} S_y^2(1 - \varrho^2)$$

Als Schätzwert für die Varianz der Regressionsschätzung erhalten wir[1)]

$$v(\bar{y}_r) = \frac{N-n}{N \cdot n} \frac{n-1}{n-2} s_y^2(1 - r^2)$$

---

1) Vgl. Sukhatme, Pandurang V. (1963): S.198ff..

Der Divisor n-2 rührt davon her, daß zwei Werte aus der Stichprobe zu schätzen sind, nämlich die Standardabweichung der y_i und der Regressionskoeffizient B. Wir haben es hier mit einer Schätzung von zwei Parametern zu tun, und die Zahl der Freiheitsgrade nimmt um einen ab gegenüber der einparametrigen Schätzung.[1]

2.3.4.3 Bias

Über den Bias wird in der Literatur nur sehr wenig gesagt.[2] Bei den meisten Autoren findet man keine[3] oder nur einige allgemeine Bemerkungen[4] zum Bias. Ein Bias tritt auf, wenn die wahre Regression zwischen x_i und y_i nicht linear ist, wir aber das Modell der linearen Regressionsschätzung zur Hochrechnung nehmen; denn dann wenden wir sozusagen das falsche Modell auf die zu untersuchende statistische Masse an. Cochran gibt als Ausdruck für den Bias an:[5]

$$B(\bar{y}_r) = \frac{N-n}{N} \frac{1}{(n-1)S_x^2} \left[\frac{\Sigma e_i (x_i - \bar{X})^2}{N-1} \right]$$

"Der Ausdruck innerhalb der Klammer ist die Kovarianz zwischen e_i und $(x_i - \bar{X})^2$ der Grundgesamtheit; sie stellt den Beitrag einer quadratischen Regression zwischen y_i und x_i dar und verschwindet, wenn die Beziehung zwischen y_i und x_i linear ist. Der Bias von $\bar{y}_r$ wird bei großen Stichproben vernachlässigbar, da er von der Ordnung $1/n$ ist, während die Standardabweichung die Größenordnung $1/\sqrt{n}$ aufweist."[6]

---

1) Vgl. Fisher, Ronald A. (1956): S.136.

2) Eine Ausnahme macht hier Cochran, William G. (1939), der sein Kapitel über den Bias in der zweiten Auflage (sec. edition, sec. print. 1964) noch erweitert und präzisiert hat.

3) So bei Deming, W.E. (1960), Hansen, Morris H./Hurwitz, William N./Madow, William G. (1953a und b), Hendricks, Walter A. (1956), Sukhatme, Pandurang V. (1963), Yates, Frank (1965).

4) So Sampford, M.R. (1962): S.134.

5) Cochran, William G. (1964): S.198.

6) Cochran, William G. (1939): S.148, eigene Übersetzung.

2.3.4.4 Die Methode von Abraham Wald

Ein großer Einwand gegen die Regressionsschätzung ist die aufwendige Rechenarbeit für die Berechnung des Regressionskoeffizienten b. Eine Möglichkeit, diese zu verringern, liefert die Methode von Abraham Wald.[1] Dabei werden die Beobachtungen so geordnet, daß x eine aufsteigende Reihe ergibt:

$$x_1 \leqq x_2 \leqq \ldots \leqq x_n$$

Die n Beobachtungen werden in zwei Gruppen aufgeteilt, wovon jede $m = \frac{1}{2}n$ Paare (x_i, y_i) enthält; ist n eine ungerade Zahl, so wird das Paar (x_i, y_i) weggelassen, für welches $i = \frac{1}{2}(n+1)$. Für diese zwei Gruppen werden die Mittelwerte errechnet:

$$\bar{x}_1 = \frac{1}{m}\sum_1^m x_i \qquad \bar{y}_1 = \frac{1}{m}\sum_1^m y_i$$

$$\bar{x}_2 = \frac{1}{m}\sum_1^m x_{m+i} \qquad \bar{y}_2 = \frac{1}{m}\sum_1^m y_{m+i}$$

Walds Schätzwert für b ist dann:

$$b_w = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}$$

Als Schätzwert für a gibt er an:

$$a_w = \bar{y} - b_w\bar{x}$$

Dabei sind $\bar{x}, \bar{y}$ die Mittelwerte aus der gesamten Stichprobe:

$$\bar{x} = \frac{1}{n}\sum_1^n x_i \qquad \bar{y} = \frac{1}{n}\sum_1^n y_i$$

Es wäre nun genau zu untersuchen, wie stark die Wirksamkeit der Schätzmethode bei Anwendung dieses Schätzwertes leidet. Die Abnahme der Kosten wäre dann gegen eine Abnahme der Wirksamkeit abzuwägen.

---

[1]) S. Wald, Abraham (1940): S.284ff., und
Theil, H./Yzeren, van J. (1956): S.17ff..

2.3.4.5 Separate und kombinierte Regressionsschätzung

Genau wie bei der Verhältnisschätzung können wir die Regressionsschätzung mit einem geschichteten Auswahlverfahren kombinieren. Dabei entstehen theoretisch keine neuen Probleme. Bei der separaten Regressionsschätzung wird für jede Schicht getrennt (separat) mit Hilfe der Regressionsschätzung der Schätzwert errechnet. Für die h-te Schicht wird dann

$$\bar{y}_{rh} = \bar{y}_h + b_h(\bar{X}_h - \bar{x}_h)$$

Dabei benötigen wir die y-, x- und b-Werte für jede einzelne Schicht. Die Formel für den Schätzwert des Mittelwertes der separaten Regressionsschätzung lautet dann für L Schichten

$$\bar{y}_{rs} = \frac{\sum N_h \bar{y}_{rh}}{N}$$

Ist die Besetzung in den einzelnen Schichten gering, oder ist für alle Schichten ein gleicher Schätzwert b für den Regressionskoeffizienten anzunehmen, so empfiehlt sich die kombinierte Regressionsschätzung. Zur Schätzung berechnen wir zunächst die Schätzwerte wie beim geschichteten Auswahlverfahren (st = stratification):

$$\bar{y}_{st} = \frac{\sum N_h \bar{y}_h}{N} \qquad \text{bzw.} \qquad \bar{x}_{st} = \frac{\sum N_h \bar{x}_h}{N}$$

Die Schätzformel der kombinierten Regressionsschätzung wird dann:

$$\bar{y}_{rk} = \bar{y}_{st} + b(\bar{X} - \bar{x}_{st})$$

2.3.5 Differenzenschätzung

Bei der Differenzenschätzung wird angenommen, daß die Differenzen von Größen in einer Gesamtheit weniger variieren als die Größen selber. Die Schätzformel sieht folgendermaßen aus:

$$\bar{y}_d = \bar{y} + k(\bar{X} - \bar{x})$$

$$k = \text{const.}$$

Der Grundgedanke, auf dem diese Schätzformel beruht, besteht darin, daß zur einfachen Hochrechnung $\bar{y}$ ein Korrekturglied $k(\bar{X}-\bar{x})$ hinzugefügt wird.[1] Dieses Korrekturglied kann nur richtig wirksam werden, wenn die zusätzliche Information über das untersuchte Merkmal eine wirkliche Aussage machen kann. Aus diesem Grund muß die zusätzliche Information sorgfältig und mit Geschick ausgewählt werden. Slonim[2] spricht in diesem Zusammenhang vom "intuitive appeal" der Differenzenschätzung. Ein bedeutender Sonderfall bei der Anwendung der Differenzenschätzung ist die Schätzformel mit $k = 1$.

Die Differenzenschätzung ist eine Sonderform zwischen Verhältnisschätzung und Regressionsschätzung.[3] Die Analogie zur Verhältnisschätzung wird deutlich, wenn wir die Gleichung umformen und nach k auflösen:

$$k = \frac{\bar{y}_d - \bar{y}}{\bar{X} - \bar{x}}$$

Bezeichnen wir die Abweichungen in den Variablen y und x mit y' und x', erhalten wir für k das Verhältnis

$$k = \frac{y'}{x'}$$

Mehr noch ist die Differenzenschätzung mit der Regressionsschätzung verwandt.[4] Nehmen wir für den Regressionskoeffizienten b irgendeinen konstanten Wert k an aus einer früheren Stichprobe, oder sonst einen Erfahrungswert, dann sehen wir die Analogie zur Regressionsschätzung sehr deutlich.[5] Weil die Differenzenschätzung also nur eine Sonderform der schon behandelten Schätzmethoden darstellt, wollen wir sie hier kürzer behandeln und weisen auf die Ausführungen über diese schon behandelten Schätz-

---

1) Vgl. Arkin, Herbert (1963): S.215, und
Slonim, Morris James (1960): S.70.

2) Slonim, Morris James (1960): S.70.

3) Vgl. Menges, Günter (1959): S.143.

4) Vgl. United Nations (Hrsg.)(1960): S.43.

5) Vgl. Cochran, William G. (1964): S.190ff., und
Désabie, M.J. (1966): S.219.

methoden hin. Die Differenzenschätzung ist eine erwartungstreue Schätzmethode, wie sich leicht zeigen läßt. Ihre Varianz ist

$$V(\bar{y}_d) = \frac{N-n}{N \cdot n} (S_y^2 + k^2 S_x^2 - 2k\varrho S_x S_y)$$

Die Varianz der Differenzenschätzung wird ein Minimum, wenn k=B wird.[1] Man kann dies leicht zeigen, indem man die erste und zweite Ableitung bildet.

Genau wie bei der Verhältnisschätzung und der Regressionsschätzung ist es möglich, die Differenzenschätzung mit einem geschichteten Auswahlverfahren zu kombinieren und wieder eine separate oder kombinierte Differenzenschätzung zu bilden.

3 A-posteriori-Schichtung

3.1 Prinzip dieser Methode

Bei Stichprobenerhebungen ist es aus Genauigkeitserwägungen heraus oft ratsam, nach der Erhebung die Elemente nach bestimmten Merkmalen noch zu schichten. Diese Eigenschaft des praktisch "garantierten Gewinns an Genauigkeit" bei geschichteten Verfahren allgemein ist auch verantwortlich für deren Beliebtheit in der praktischen Stichprobenerhebung.[2] Darüber hinaus ist es aus Kosten- und Zeitersparnisgründen äußerst nützlich und in sehr vielen Fällen, nämlich bei Faktoren wie Einkommen, Beschäftigung, politischer Einstellung, Religionszugehörigkeit usw., die einzige Möglichkeit, die Schichtung nur bei den Stichprobendaten, d.h. nach der Erhebung vorzunehmen.

---

1) Vgl. Dalenius, Tore (1957): S.20.

2) Bryant, Edward C./Hartley, H.O./Jessen, R.J. (1960): S.105.

In allen diesen Fällen: aus Genauigkeits-, Kosten- und Zeitersparnisgründen, oder wenn die Kenntnis einer Merkmalszugehörigkeit erst nach der Befragung offenkundig wird, ist es "vollkommen legitim, die per Zufall ausgewählten Elemente nach der Auswahl zu schichten".[1] Die Methode dieser Hochrechnungsart ist folgende[2]: Sobald die Daten vorliegen, bilden wir mit Hilfe bestimmter Charakteristika verschiedene Schichten, auf die die Stichprobeneinheiten verteilt werden. Die Schichten müssen dafür eindeutig definiert und abgegrenzt sein, damit eine Einheit nur jeweils einer Schicht zugerechnet werden kann. Die Mittelwerte $\bar{Y}_1, \bar{Y}_2, \ldots \bar{Y}_h, \ldots \bar{Y}_L$ jeder Schicht werden einzeln durch die Mittelwerte der Stichprobe $\bar{y}_1, \bar{y}_2, \ldots \bar{y}_h, \ldots \bar{y}_L$ geschätzt. Der Mittelwert der Grundgesamtheit $\bar{Y}$ wird dann geschätzt mit:

$$\bar{y}_p = \sum \frac{N_h}{N} \bar{y}_h = \frac{1}{N} \sum \frac{N_h}{n_h} \sum_i y_{hi}$$

wobei N_h bzw. n_h die Anzahl der Einheiten in den Schichten der Grundgesamtheit bzw. Stichprobe angeben. N ist die Gesamtzahl der Einheiten in der Grundgesamtheit.

Wie wir sehen, benützen wir die gleiche Schätzformel wie bei der A-priori-Schichtung (geschichtetes Auswahlverfahren) bei proportionaler Aufteilung. Man kann jedoch weder zwischen den Schichten eine optimale Aufteilung vornehmen, noch für alle Schichten einen einheitlichen Erhebungssatz festlegen wie bei jener.[3] Die Varianzformel, die wir an späterer Stelle behandeln wollen, wird jedoch eine andere Form aufweisen.

Diese nachträglich geschichtete Zufallsstichprobe ist verwandt mit der A-priori-Schichtung (geschichtetes Auswahlverfahren), wie man sofort sieht. Es werden bei der A-posteriori-Schichtung etwa ähnliche Genauigkeitsgewinne erzielt wie bei dieser.[4] Ein wesentlicher Unterschied zwischen beiden Verfahren besteht

---

[1]) Sampford, M.R. (1962): S.98.
[2]) Vgl. Désabie, J. (1966): S.174f..
[3]) Vgl. Désabie, J. (1966): S.174.
[4]) Vgl. Sukhatme, Pandurang V. (1963): S.107/108.

jedoch darin, daß

- die A-priori-Schichtung ein Auswahlverfahren, die A-posteriori-Schichtung jedoch ein Hochrechnungsverfahren ist (s. folgender Abschnitt 3.2 Die A-posteriori-Schichtung, ein Post-hoc-Verfahren und seine Vorteile), und daß
- bei der A-posteriori-Schichtung die Schichtgröße in der Stichprobe nicht von vornherein bestimmt werden kann; sie ist eine Zufallsgröße.

Natürlich könnte mit der Auswahl solange fortgefahren werden, bis eine bestimmte, vorher fixierte Anzahl von Einheiten in jeder Schicht ausgewählt ist. Dabei werden dann die überzähligen Beobachtungen ignoriert.[1] Dieses Verfahren wäre wesentlich kostspieliger als die A-posteriori-Schichtung, und es käme der A-priori-Schichtung mit proportionaler Auswahl, bzw. beim Einsatz von Interviewern, dem "randomisierten" Quotenverfahren gleich.[2] (s. Abschnitt 3.3.2 A-posteriori-Schichtung und Quotenverfahren)

Wir haben also gesehen, daß wir nach erfolgter Auswahl, wenn die Stichprobenergebnisse vorliegen, die Stichprobenelemente in Schichten einteilen. Daraus berechnen wir für jede Schicht einen Mittelwert. Nun benötigen wir zur Hochrechnung noch die richtigen Anteilswerte der Schichten, um die Schichten richtig gewichten zu können. Am einfachsten liegt der Fall, wenn wir die tatsächlichen Gewichte N_h/N der Grundgesamtheit zur Verfügung haben. Es tritt dann in den Gewichten weder ein Bias auf, noch stellt sich das Problem, die Gewichte schätzen zu müssen. Mit beiden Problemen - Verzerrung in den Gewichten und Schätzung der Gewichte - werden wir uns noch zu befassen haben. Nach der einfachen Berechnung der Schichtmittelwerte bilden also die Gewichte der einzelnen Schichten die zusätzliche Information für die Hochrechnung. Bei dieser Methode zur Gewinnung guter Schätzungen aus Stichproben können wir also von einer A-poste-

---

1) S. Williams, W.H. (1962): S.622.

2) Vgl. Cochran, William G. (1964): S.130f. und Kish, Leslie (1965): S.99.

riori-Schichtung reden oder genauer, vom "Gebrauch einer Schätzformel für geschichtete Stichproben", im Gegensatz zur A-priori-Schichtung, bzw. vom "Ziehen einer Stichprobe aus einer geschichteten Grundgesamtheit".[1)]

3.2 Die A-posteriori-Schichtung, ein Post-hoc-Verfahren und seine Vorteile

In der statistischen Theorie, aber auch in der statistischen Betriebslehre, werden eine Reihe von Methoden angegeben, um die Qualität statistischer Angaben zu verbessern. Fehler können verkleinert, bzw. beseitigt werden, z.B. durch: Vergleich der Daten mit entsprechenden Daten aus früherer Zeit, bzw. von anderen Institutionen[2)], Konsistenzprüfung, Beobachtung einer bestimmten (z.B. Alters-) Gruppe über mehrere Zählungen hinweg (study of cohort survival), Listenvergleich usw..[3)] Diese Verbesserungen sind größtenteils technischer Art, also Probleme, die in der statistischen Betriebslehre behandelt werden. Hier können keine Vertrauensgrenzen angegeben werden wie bei der Zufallsstichprobe. Darüber hinaus ergibt sich der Nachteil, daß die verschiedenen Typen von Fehlern nicht mehr voneinander getrennt werden können, da diese Datenverbesserungen sich auf den Gesamtirrtum beziehen.[4)] Man versucht zwar, eine möglichst vollständige Information zu erhalten, dabei vergißt man aber, daß es im Wesen der Stichprobe liegt, unvollständige Informationen zu liefern; diese Informationen werden aber - und das ist das Charakteristikum der Stichproben - mit Vertrauensintervallen geliefert.

---

1) Thionet, M.P. (1953a): S.239, eigene Übersetzung.

2) Dabei ist aber sehr genau auf die gleichen Definitionen zu achten.

3) Diese Probleme und Methoden werden sehr ausführlich behandelt bei Zarkovich, S.S. (1964): S.41 und S.142.

4) S. Zarkovich, S.S. (1964): S.374.

Aus diesem Grund versucht man im Stadium der Extrapolation der Daten, d.h. bei der Hochrechnung, Schätzformeln zu finden, die die Zufallsfluktuation vermindern und genauere Schätzwerte liefern, d.h. engere Vertrauensintervalle angeben.[1] Daß diese Schätzformeln dann wesentlich komplizierter sind, versteht sich von selbst, da in diesen mehrere Informationen - mindestens aber eine zusätzliche Information neben der untersuchten Stichprobe - zur Gewinnung der Schätzwerte verarbeitet werden. Solche Überlegungen führten zu den gebundenen Hochrechnungsarten, und zu diesen gehört auch die A-posteriori-Schichtung, denn

- auch sie liefert uns genauere Ergebnisse als die einfache Hochrechnung, weil sie eine gebundene Hochrechnung ist, und
- sie kommt erst zur Anwendung, wenn die Daten schon vorliegen, d.h. der Effekt der Schichtung - im Gegensatz zur A-priori-Schichtung - wird in den Hochrechnungsprozess eingegliedert.[2]

Im Gegensatz zu den anderen bekannten Hochrechnungsarten (s. Abschnitt 2.3) stellt man über die Beziehung der Variablen y und x (untersuchte Größe und zusätzliche Information) keine Überlegungen derart an, ob diese Beziehung in einem Verhältnis, einer Differenz oder einer Regression besteht. Man kann beinahe sagen, das Ziel der Schichtung liegt in ihr selbst.[3] Die zusätzliche Information gibt nur an, wie die untersuchte Variable geschichtet werden soll und wie man zu homogenen Schichten gelangt. Da also keine Modellvorstellung (wie z.B. das Regressionsmodell bei der Regressionsschätzung) zugrunde liegt, kann auch kein falsches Modell angewandt werden, wie dies bei der Regressionsschätzung möglich ist, wenn wir eine Gerade statt, wie es notwendig wäre, eine Parabel als Regressionskurve zugrunde legen.

Ferner werden gewöhnlich Verhältnis-, Regressions- oder Differenzenschätzung nicht miteinander kombiniert. Wohl kann aber die A-posteriori-Schichtung mit einem dieser anderen Hochrechnungs-

---

1) Vgl. Désabie, J. (1966): S.173.
2) Vgl. Kish, Leslie (1965): S.90.
3) Vgl. United Nations (Hrsg.)(1960): S.37.

verfahren gekoppelt werden. Die A-posteriori-Schichtung ermöglicht also eine Kombination zweier Hochrechnungsverfahren (s. Abschnitt 3.8).

Auf die Verwandtschaft der A-posteriori-Schichtung mit der A-priori-Schichtung haben wir schon hingewiesen.[1] Aber, daß es eine Frage der Bequemlichkeit sei, welche von beiden Methoden angewandt werde[2], scheint mir nicht einleuchtend; denn

- die A-posteriori-Schichtung ist kein Auswahlverfahren, sondern ein Hochrechnungsverfahren,
- die Schichtgröβe ist nicht fixiert wie bei der A-priori-Schichtung, sondern variabel,
- es wird nicht die Grundgesamtheit geschichtet, sondern nur die Stichprobe,
- die Schichtung bei der A-priori-Schichtung kann nur nach altem Material vorgenommen werden (vor der Auswahl); bei der A-posteriori-Schichtung kann dies prinzipiell immer nach neuestem Material bzw. neuesten zusätzlichen Informationen vorgenommen werden.

Die Vorteile der A-posteriori-Schichtung wurden zum Teil schon angedeutet; sie sollen hier noch einmal klar herausgestellt werden:

- Die A-posteriori-Schichtung ist ein sehr wirksames Hochrechnungsverfahren (Der theoretische Teil ebenso wie die Beispiele an späterer Stelle werden dies zeigen.). Sie erbringt zum groβen Teil die Gewinne der Stichprobe mit proportional geschichtetem Auswahlverfahren.
- Sie ist ein arbeits-[3] und somit kosten- und zeitsparendes Verfahren. Sie stellt keinen Mehraufwand für die Vorbereitung (wie z.B. die A-priori-Schichtung) noch für die Feldarbeit dar. Bei der Hochrechnung ist es bei elektronischen Rechenanlagen relativ einfach, genügend viele Schichten zu bilden und einen möglichst guten

---

1) Ebenso: Cochran, William G. (1964): S.135f.; Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a): S.232 und (1956b): S.138f.; Sampford, M.R. (1962): S.98; Sukhatme, Pandurang V. (1963): S.106ff..

2) Yates, Frank (1965): S.25.

3) Statistisches Bundesamt (Hrsg.)(1960): S.48/49.

Schichtungseffekt zu sichern.

- Bei schon vorhandener Schichtung können noch nachträglich weitere (nach anderen Merkmalen gebildete) Schichten eingeführt werden.
- Oft ist eine nachträgliche Schichtung die einzige Möglichkeit (politische Einstellung, Religionszugehörigkeit usw.).[1)]
- Das große Problem bei Stichprobenuntersuchungen ist die strukturelle Veränderung der Grundgesamtheit in der Zeit.[2)] So wird bei geschichteten Auswahlverfahren die Auswahl nach altem Material oder/und Hypothesen vorgenommen. Dadurch kann ein Bias eingeführt werden, der selbst optimale Schichtaufteilung nur angenähert, bzw. nur wenn keine strukturellen Veränderungen auftreten, zuläßt. Das hat auch eine Arbeit im Institut für Wirtschafts- und Sozialstatistik Tübingen über Stichprobenerhebungen bei Viehzählungen in Belgien (im Rahmen der EWG) gezeigt, wo deutlich wurde, daß Schichtenbildung in der Grundgesamtheit - von dem vorliegenden Material ausgehend, das meist veraltet war - wegen laufender Strukturänderungen Schwierigkeiten mit sich bringt.
- Die A-posteriori-Schichtung läßt sich auch mit anderen Hochrechnungsarten zur Verbesserung der Wirksamkeit kombinieren.

Als Erfordernisse müssen natürlich vorhanden sein:

- die Schichtaufteilung als zusätzliche Information,
- die Schichten müssen besetzt sein, aber dabei kann der Untersuchende bei der Wahl der Schichten bis zuletzt variabel bleiben; er behält sozusagen die Konzeptionswahl bis zum Schluß. Jedoch wird auch für diesen Fall (Nichtbesetzung von Schichten) eine Lösung gebracht (s. Abschnitt 3.4.2).

---

1) Vgl. Kish, Leslie (1965): S.99.
2) Vgl. United Nations (Hrsg.)(1960): S.5.

3.3 A-posteriori-Schichtung und verwandte Verfahren

Mit einigen Verfahren hat die A-posteriori-Schichtung Berührungspunkte. Das Gemeinsame und Trennende dieser Verfahren - Zwei-Phasen-Stichprobe, Post-Klumpenstichprobe (post-cluster-sampling), Quotenverfahren und Gruppierungsverfahren (grouping) - wollen wir in den folgenden Abschnitten aufzeigen.

3.3.1 Zwei-Phasen-Stichprobe und Post-Klumpenstichprobe

Die Mehr-Phasen-(und speziell die Zwei-Phasen-)Stichprobe ist wie die A-posteriori-Schichtung ein Post-hoc-Verfahren. Das Problem wurde zuerst 1937 von Friedman und Wilcox auf einer Statistikerkonferenz an Neyman herangetragen[1]. Zur Verbesserung der Ergebnisse, um zusätzlich Ergebnisse über andere als die untersuchten Merkmale zu erhalten, oder wenn es keine andere Möglichkeit gibt, wie z.B. in Deutschland 1945[2], zieht man die Stichprobe nicht aus der Grundgesamtheit, sondern ebenfalls aus einer Stichprobe. Das Prinzip ist also, daß eine Stichprobe aus einer Stichprobe gezogen wird, um - ganz allgemein - (zusätzliche) Informationen zu erhalten[3]. Man kann aber auch mit Hilfe der Zwei-Phasen-Stichprobe - und hier gewinnt sie an Bedeutung für die A-posteriori-Schichtung - die Schichtanteile für eine (nachträglich) geschichtete Stichprobe schätzen[4]. Auf dieses Problem wollen wir in einem späteren Abschnitt über die Schichtenbildung eingehen.

Der wesentliche Unterschied zwischen beiden Verfahren ist der, daß die Phasenstichprobe ein (nachträgliches) Auswahlverfahren, wogegen die A-posteriori-Schichtung ein Hochrechnungsverfahren ist.

---

1) S. Ghosh, Sakti P. (1963a): S.587.
2) Vgl. Cochran, William G. (1964): S.327.
3) Vgl. Kish, Leslie (1965): S.327.
4) Vgl. Sukhatme, Pandurang V. (1963): S.112ff..

Auf dem gleichen Prinzip wie die Phasen-Stichprobe beruht die Post-Klumpen-Stichprobe[1]. Auch sie ist - im Gegensatz zur A-posteriori-Schichtung - ein nachträgliches Auswahlverfahren, wobei aus einer vorliegenden Stichprobe eine Klumpenstichprobe gezogen wird. Sie ist vor allem dann zweckmäßigerweise anzuwenden, wenn die Zusammensetzung der Klumpen von vornherein nicht bekannt oder nur sehr schwer zu ermitteln ist.

Die Post-Klumpen-Stichprobe ähnelt sehr dem Quotenverfahren - das wir im folgenden Abschnitt behandeln wollen -, wobei jedoch ersteré auf einem stochastischen Modell beruht und somit Vertrauensintervalle angegeben werden können, was normalerweise für das Quotenverfahren nicht gilt.

3.3.2 A-posteriori-Schichtung und Quotenverfahren

Wie wir in Abschnitt 3.1 (Prinzip dieser Methode) gezeigt haben, besitzen A-posteriori-Schichtung und Quotenverfahren gemeinsame Züge, ja sie können in Grenzfällen ineinander übergehen. Für lange Zeit war das Quotenverfahren das vorherrschende Stichprobenverfahren bei Meinungsumfragen[2]. Auch der Nichtstatistiker assoziiert beim Namen "Gallup" den Begriff "Meinungsumfrage" und umgekehrt. Die Qualität des Quotenverfahrens war und ist noch heute umstritten. Dabei wird oft der Schluß gezogen, das Quotenverfahren sei billiger als die Zufalls- bzw. Wahrscheinlichkeitsstichprobe, also müsse es auch ungenauer sein.[3] Dieses Vorurteil scheint nicht nur auf den ersten Blick voreingenommen und oberflächlich.

Es ist schwierig, genau zu sagen, was das Quotenverfahren ist[4]. Allgemein gilt: Beim Quotenverfahren wird die Grundgesamtheit in sich ausschließende und erschöpfende Untergruppen aufgeteilt.

---

1) S. den Artikel von Ghosh, Sakti P. (1963a): S.587-597, in dem dieses Verfahren ausführlich bei gleicher und ungleicher Auswahlwahrscheinlichkeit behandelt wird.

2) S. Stephan, Frederick F./McCarthy, Philip J. (1958): S.37.

3) S. Sudman, Seymour (1966): S.749.

4) Vgl. Kish, Leslie (1965): S.562.

Z.B. wird die Bevölkerung mit 20 Jahren und mehr der Bevölkerungsstruktur nach aufgeteilt in folgende Gruppen

(a) in Ausbildung sich befindend
(b) berufs- oder erwerbstätig
(c) nicht oder nicht mehr berufs- oder erwerbstätig.

Die Daten dazu erhält man aus der Volkszählung oder anderen Quellen (z.B. Mikrozensus). Dem Interviewer wird nun eine bestimmte Quote jeder Untergruppe zugeteilt, die er zu befragen hat, und die er auf einer ihm vorgeschriebenen Reiseroute dann auszuwählen hat. Welche Personen der Interviewer in einer bestimmten Untergruppe dann individuell auswählt, bleibt ihm überlassen. In der Agrarstatistik im besonderen und in der amtlichen Statistik im allgemeinen ist das Quotenverfahren von geringer Bedeutung.[1)]

Die Vorteile des Quotenverfahrens sind:

- relativ große Erfolge in der Vergangenheit,
- großer Anwendungsbereich,
- kostensparend,
- große Schnelligkeit bei der Lieferung von Ergebnissen.

Die bekanntesten Nachteile sind[2)]:

- ab und zu große Mißerfolge, z.B. bei der amerikanischen Präsidentenwahl 1948 oder bei der Wahl zum deutschen Bundestag 1965, wo für CDU und SPD ein Kopf-an-Kopf-Rennen vorausgesagt wurde, die Wahl jedoch mit Abstand von der CDU gewonnen wurde,
- das Quotenverfahren ist eine Variante der bewußten Auswahl; es ist nicht äquivalent mit der Zufallsstichprobe,
- im Hintergrund steht ein Bias, der hauptsächlich durch den Interviewer verursacht werden kann,
- über die Vertrauensgrenzen ist keine Angabe möglich.

Die letzten Gründe betreffen das gleiche Problem, nämlich die Auswahl durch den Interviewer innerhalb der Quoten. Diese Auswahl ist willkürlich oder bewußt, auf jeden Fall nicht zufalls-

---

1) Vgl. Sampford, M.R. (1962): S.XVII; Statistisches Bundesamt (Hrsg.)(1960): S.30, wo bemerkt wird, daß in der amtlichen Statistik in Deutschland das Quotenverfahren nur gelegentlich bei der besonderen Ernteermittlung vorkommt.

2) Vgl. Yates, Frank (1965): S.80.

bedingt; sie wird vom individuellen Urteil des Interviewers gelenkt. Oft wird auch zugegeben, daß Interviewer Reisekosten und -mühen sparen, indem sie nur den 1.Stock eines Hauses besuchen, oder Parks "abgrasen", wo Mütter auf ihre Kinder aufpassen[1]. Sehr umständlich und etwas erheiternd wirken die Bemühungen der Meinungsforschungsinstitute, diesen Schwierigkeiten zu begegnen. So kann man in der Random-Route-Anweisung des DIVO-Institutes unter anderem lesen: "Zählen Sie - ausgehend von der vorgegebenen Originaladresse - stets zwei Haushalte ab, und zwar wenden Sie sich nach rechts (mit dem Gesicht zum Ausgangshaushalt) und beginnen mit dem Abzählen. Dabei bleiben die Haushalte, die zwar im selben Stockwerk, jedoch links vom Ausgangshaushalt liegen, unberücksichtigt. Sind im selben Stockwerk keine weiteren Haushalte oder nicht genügend rechtsgelegene Haushalte vorhanden, so gehen Sie bitte in das darüberliegende Stockwerk und zählen die beiden Haushalte dort im Uhrzeigersinn - das heißt von links nach rechts gehend - ab. Gibt es im selben Hause keine weiteren Stockwerke beziehungsweise weiteren Haushalte, dann setzen Sie die Zählung im Nachbarhaus fort, beginnend im Erdgeschoß. Das Nachbarhaus finden Sie, indem Sie - mit dem Gesicht zur Originaladresse - nach rechts gehend ...."[2]. Der Versuch, die Interviewer zu kontrollieren, indem diese die Namen der Interviewten aufschreiben und damit eine Nachkontrolle ermöglichen, schlägt fehl; denn dies führt zu einem Eingriff in die Anonymität des Befragten, der dann schlechte Ergebnisse erwarten läßt.[3] Dabei ist natürlich immer zu beachten, auch wenn die Anonymität strikt gewahrt bleibt, daß die Aussagekraft der Befragten ein Problem für sich darstellt. Dies gehört in das Gebiet der Konsistenzforschung.[4]

Die Quoten stehen für Schichten, in die die Grundgesamtheit aufgeteilt wurde. Die Elemente oder Individuen, die in eine Schicht gelangen - und damit haben wir die wichtigste Gemeinsamkeit mit

---

1) Vgl. Kish, Leslie (1965): S.564.

2) So unter dem Titel "Nonsens im Uhrzeigersinn", nachzulesen in der Süddeutschen Zeitung vom 16./17. Dezember 1967, S.68.

3) Vgl. Noelle, Elisabeth (1963): S.148.

4) S. Stroschein, Fritz-Reinhard (1965): S.16.

der A-posteriori-Schichtung - kennt man erst nach der Befragung. Würden die Elemente per Zufall ausgewählt, hätten wir ein geschichtetes Stichprobenverfahren, wobei aber die Schichten in der Stichprobe, die Quoten, als Blankett vom Interviewer ausgefüllt werden[1]. Im Unterschied zur A-posteriori-Schichtung ist die Schichtgröße dabei von vornherein festgelegt und nicht variabel. Da wir beim Quotenverfahren kein Vertrauensintervall zur Wirksamkeitskontrolle und zum Wirksamkeitsvergleich angeben können, da ferner neben dem Problem der Quotenaufteilung und Schichtendefinition das subjektive Element bei der Auswahl der Befragten durch den Interviewer eine äußerst wichtige Rolle spielt, ist es sehr schwer von dem Quotenverfahren zu sprechen. Zwei Quotenverfahren, von denen man annimmt, sie seien einander ähnlich, müssen aufs genaueste beschrieben werden, erst dann kann man sagen, "die zwei Verfahren sind einander etwa ähnlich"[2]. Man kann also nicht einfach vom Quotenverfahren sprechen wie etwa von der Zufallsstichprobe, sondern das Verfahren muß genau beschrieben werden, um es in seiner Genauigkeit würdigen zu können. Obwohl das Ideal der Quotenauswahl die Zufallsstichprobe ist[3], erreicht sie dieses Ideal nicht, weil sie das ausschlaggebende Charakteristikum der Zufallsstichprobe und somit auch der A-posteriori-Schichtung nicht besitzt, nämlich die Angabe von Vertrauensgrenzen.

3.3.3 A-posteriori-Schichtung und Gruppierungsverfahren (grouping)[4]

Das "Grouping"-Problem umfaßt zwei Problemkreise. Einmal den einfachen Vorgang, daß man der bequemeren Rechenweise halber die Daten gruppiert. Man erhält dabei einen Gruppenfehler, denn

---

1) Vgl. Kish, Leslie (1965): S.99.

2) Stephan, Frederick F./McCarthy, Philip J. (1958): S.38.

3) Noelle, Elisabeth (1963): S.148.

4) Da der deutsche Ausdruck einmal schwerfällig, zum anderen nicht besonders gebräuchlich ist, wird hier öfters der englische Ausdruck "grouping" gebraucht.

"grouping" ist immer mit einem Verlust an Genauigkeit verbunden[1], den man mit Hilfe eines Korrekturfaktors berichtigt. Nehmen wir zur Berechnung der Parameter der Gesamtheit das arithmetische Mittel, erhalten wir als Gesamtvarianz σ^2 den Ausdruck:

$$\sigma^2 = \Sigma P_i \sigma_i^2 + \Sigma P_i (\mu_i - \bar{\mu})^2$$

dabei ist: P_i der Anteilssatz jeder Gruppe
σ_i^2 die Varianz jeder Gruppe
μ_i das arithmetische Mittel jeder Gruppe
$\bar{\mu}$ das arithmetische Mittel über alle Gruppen

Der erste Summenausdruck ist die Varianz der Gruppen zusammen, der zweite Summenausdruck ist der Korrekturfaktor, der bei gleichem Mittelwert gleich 0 wird. Nehmen wir nun statt des arithmetischen Mittels jeder Gruppe die Gruppenmitte (Klassenmitte), so muß die summierte Varianz der Gruppen normalerweise um einen Faktor δ korrigiert werden, da man annimmt, die Häufigkeiten seien in der Klassenmitte konzentriert. Wenn $\sigma_g^2 = \Sigma P_i \sigma_i^2$ die gewichtete Varianz der Gruppen ist, dann erhalten wir für die korrigierte Gesamtvarianz den Ausdruck[2]:

$$\sigma^2 = \sigma_g^2 \pm \delta$$

Bei gleichgroßen Klassenintervallen a erhalten wir unter bestimmten Voraussetzungen, wovon die wichtigste die Annahme einer unimodalen Verteilung ist, d.h. einer Verteilung, die in beiden Richtungen gegen 0 tendiert[3], als Korrekturfaktor $\delta = -a^2/12$. Dieser Korrekturfaktor ist der bekannteste und gehört in die Reihe der Sheppard'schen Korrekturen[4]. Die Anwendung dieses

---

1) Vgl. Yates, Frank (1965): S.239.

2) Vgl. Yule, Udny G. (1927): S.211ff..

3) Diese Voraussetzung wird nicht immer gemacht, z.B. auch nicht bei: Van der Waerden, B.L. (1957): S.80ff., wobei dieser aber im angeführten Anwendungsbeispiel eine Normalverteilung zugrunde legt.

4) S. Lorenz, Paul (1951): S.179ff., der sich mit verschiedenen Herleitungen auseinandersetzt.

Korrekturfaktors ist jedoch umstritten, weil man annimmt, daß er zu einer Überkorrektur neigt und somit einen neuen Fehler anstelle des alten setzt[1]. Deshalb wird auch geraten, ihn erst bei n>500 anzuwenden[2]. Dieses Problem ist jedoch kein spezielles Stichprobenproblem.

Der andere Problemkreis betrifft die Gruppenbildung zur besseren Analyse, bzw. wird durch diese erst eine Analyse möglich gemacht. Diese Klassifikation bedeutet ein Ordnen von Merkmalen oder Organismen in Gruppen oder Reihen auf der Basis ihrer Beziehungen zueinander. Diese Beziehungen können in gemeinsamen Berührungspunkten und/oder Überschneidungen und/oder Ähnlichkeit der Merkmale bestehen[3]. Die Gruppenbildung zur besseren Überschaubarkeit, Analyse und Berechnung wird in allen Bereichen der Wissenschaft, vor allem aber in der Natur- und auch in der Wirtschaftswissenschaft angewandt. Zum Beispiel kann man Leontief-Modelle nicht mit Einzelunternehmen aufstellen, weil diese Modelle nicht mehr zu berechnen sind. Deshalb faßt man die Unternehmen in Industriebereiche zusammen, die aber möglichst homogen sein müssen[4]. Unter der trivialen Erkenntnis, daß wir bei einer Gruppe mit n Gliedern die größte Information erhalten, wenn wir n Gruppen bilden, bieten sich zwei Wege des Vorgehens an:
Einmal können wir mit Hilfe von Tests (χ^2-Test, Varianzanalyse) prüfen, ob eine Gruppe homogen ist. Wenn die Homogenitätsbedingung nicht erfüllt ist, bilden wir zwei Gruppen, dann womöglich drei usw.[5]. Das heißt also, wir nähern uns mit Hilfe von Probieren der besten Lösung.
Zum andern können wir auch eine Funktion suchen, die den Verlust minimiert[6]. In diesem Sinne kann man eine Lösung mit Hilfe der Abraham-Wald'schen Verlustfunktion (loss-function) erzielen[7]. Dabei ist zu beachten, ob wir es mit einem unbeschränkten

---

1) Vgl. Spiegel, Murray R. (1961): S.72.
2) S. Griffin, John I. (1962): S.110.
3) Vgl. Sokal, Robert R./Sneath, Peter H.A. (1963): S.3ff..
4) Vgl. Fisher, Walter D. (1953): S.567.
5) Vgl. Lazar, Philippe (1966): S.39-43.
6) Vgl. Cox, D.R. (1957): S.543.
7) S. Fisher, Walter D. (1953): S.568ff..

Problem, d.h. ohne irgendwelche Nebenbedingung, oder mit einem beschränkten Problem, d.h. mit Nebenbedingung (z.B. Einhaltung eines Kostenvoranschlags) zu tun haben[1].

Betrachten wir anhand eines Beispiels die unbeschränkte Problemlösung[2]. Als Funktion, die es zu minimieren gilt, nehmen wir einfach die Varianz. Wir haben folgende Daten vorliegen:

2,6,5,6,2,0,2,2,0,0.

Diese Datengruppe gibt uns als Mittelwert $\mu = 2,5$ an und für die Varianz $\sigma^2 = 5,05$. Die Varianz gibt uns also einen quantitativen Wert für den Verlust an Information an. Aus dieser einen Gruppe bilden wir nun vier Gruppen der folgenden Formation:

|0,0,0| |2,2,2,2| |5| |6,6|

und erhalten vier Varianzausdrücke mit 0. Das heißt, die Verlustfunktion (Varianz) ist bei vier Gruppen minimiert. Wir erhalten also hier unter unserer Voraussetzung ein Maximum an Informationen. Im ungünstigsten Fall müssen wir jedoch, um diese maximale Information zu erhalten, wie wir schon gesagt haben, bei einer Gruppe von n Gliedern n Gruppen bilden.

Ohne auf die unterschiedlichen Aufgabenbereiche von "grouping" und Schichtung allgemein eingehen zu wollen, besteht der Hauptunterschied zwischen beiden Verfahren darin, daß das "Grouping"-Problem kein spezielles Stichprobenproblem ist, und nur als Schichtungsproblem - ob bei A-priori- oder A-posteriori-Schichtung - in den Bereich der Stichprobentheorie gehört. Beide Problemkreise überschneiden einander, wenn die Schichtgrenzen bzw. Gruppengrenzen so gewählt werden, daß bei einem gegebenen Wert k die Varianz ein Minimum wird. Wir berühren dann die Probleme der optimalen Schichtung wie sie bei Dalenius und anderen behandelt werden.[3]

Die Gemeinsamkeiten speziell mit der A-posteriori-Schichtung sind groß. Bei beiden Methoden betrachten wir vorliegende Daten und versuchen dann homogene Gruppen zu bilden, um eine möglichst große Information zu erhalten. Das Gruppierungsproblem können

---

1) Vgl. Fisher, Walter D. (1958): S.789ff..

2) Aus Ward, Joe H. (1963): S.237.

3) Vgl. hierzu Dalenius, Tore (1950): S.203-213; Kulldorf, Gunnar (1961): S.20; Råde, Lennart (1964): S.56-69; Thionet, P. (1967): S.24.

wir als ein Spezialproblem der A-posteriori-Schichtung betrachten, wir können Lösungsvorschläge aus der "Grouping"-Theorie auch zur Lösung der A-posteriori-Schichtung verwenden.

3.4 Theorie

Unter der Theorie eines Schätzverfahrens wollen wir die theoretische Behandlung folgender Parameter verstehen:

Erwartungswert
Varianz und deren Schätzwert
Bias

3.4.1 Erwartungstreue des Schätzwerts (bei besetzten Schichten)

Nehmen wir an, die Stichprobendaten würden uns für das Merkmal y folgende Häufigkeitsverteilung liefern (f_i = absol. Häufigkeit):

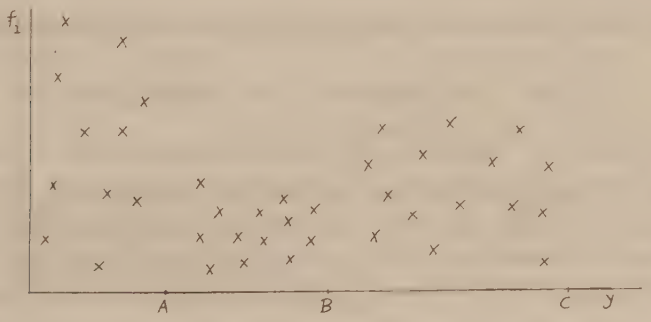


Abb.10 Häufigkeitsverteilung bei einer Zufallsstichprobe

Diese Häufigkeitsverteilung können wir, wie aus Abbild 10 ersichtlich ist, in den Punkten A,B und C in Schichten einteilen. In jeder Schicht werden Mittelwert und Varianz berechnet und diese mit Hilfe der entsprechenden Gewichte zu den Parametern der gesamten Häufigkeitsverteilung zusammengefaßt, die wir

letztlich suchen.

Auf diesem Prinzip, nämlich, daß (auch) die Stichprobendaten in Schichten eingeteilt werden können (nicht nur die Grundgesamtheit wie beim geschichteten Auswahlverfahren), beruht die A-posteriori-Schichtung.

Folgenden Mittelwert wollen wir als Parameter der Grundgesamtheit schätzen (bei h=1,...,L Schichten):

$$\bar{Y} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{Y}_h$$

Dafür nehmen wir bei der A-posteriori-Schichtung folgende Schätzformel:

$$\bar{y}_p = \sum_{h}^{L} \frac{N_h}{N} \bar{y}_h = \frac{1}{N} \sum_{h}^{L} \frac{N_h}{n_h} \sum_{i}^{n_h} y_{hi} \qquad (1)$$

Dabei ist:

L	die Anzahl der Schichten
y_{hi}	der Wert der i-ten Einheit in der h-ten Schicht
n_h	die Zahl der Stichprobeneinheiten in der h-ten Schicht
$\bar{y}_h$	das arithmetische Mittel in der h-ten Schicht
N_h	die Anzahl der Grundgesamtheitseinheiten in der h-ten Schicht
N	die Zahl der Elemente in der Grundgesamtheit

N_h/N bildet also das Gewicht, das uns als zusätzliche Information zu den Stichprobendaten geliefert werden muß. Der Fall, daß die Gewichte nicht aus der untersuchten Grundgesamtheit sind, sondern aus anderer Quelle oder über hypothesen zur Anwendung gelangen, wollen wir an späterer Stelle behandeln.

Der Beweis für die Erwartungstreue des Schätzwerts der A-posteriori-Schichtung wird genauso geführt wie bei dem geschichteten Auswahlverfahren (A-priori-Schichtung). Die Voraussetzung ist dabei, daß alle Schichten besetzt sind, d.h. n_h muß ungleich 0 sein[1]. Diese Voraussetzung gilt natürlich auch prinzipiell

---

[1]) Vgl. Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956b): S.138.

für die A-priori-Schichtung. Wenn es sich herausstellt, daß nach der Auswahl eine Schicht aus irgendeinem Grund nicht besetzt sein sollte, liefert das Verfahren verzerrte Schätzwerte. Bei der A-priori-Schichtung ist dann das gesamte Verfahren hinfällig, wenigstens in bezug auf die Erwartungstreue. Dagegen können wir bei der A-posteriori-Schichtung die Schätzformel anwenden, die wir im folgenden Abschnitt (3.4.2) behandeln, oder aber wir versuchen mit Hilfe anderer Gewichte eine andere Schichteinteilung vorzunehmen, da wir ja bei einem Hochrechnungsverfahren bis zuletzt freie Hand bei der Konzeptwahl haben. Nur ist dann zu vermuten, wenn wir sozusagen ein zweitbestes Verfahren wählen, daß wir weniger wirksame Schätzwerte erhalten.

Der Beweis der Erwartungstreue ist sehr einfach, sofern angenommen werden kann, daß der Schätzwert für jede Schicht unverzerrt ist. Dies wird allgemein angenommen bei einem Stichprobensprich Schichtumfang von $n_h > 20$.[1)]

Dann gilt:

$$E(\bar{y}_p) = E\left[\sum_h^L \frac{N_h}{N} \bar{y}_h\right] = \sum_h^L \frac{N_h}{N} E(\bar{y}_h)$$

$$= \sum_h^L \frac{N_h}{N} \cdot \bar{Y}_h = \bar{Y}$$

Daraus ergibt sich sofort der Schätzwert für den Totalwert Y der Grundgesamtheit:

$$\hat{Y}_p = N\bar{y}_p = \sum_h^L N_h \bar{y}_h \qquad (2)$$

Im folgenden Abschnitt wollen wir den allgemeinen Fall behandeln, d.h. den Erwartungswert des Schätzwerts bilden, auch wenn eine Schicht nicht besetzt ist.

---

1) Vgl. Cochran, William G. (1964): S.135; ebenso Sasieni, Maurice/Yaspan, Arthur/Friedman, Lawrence (1959): S.44. Dagegen spricht O'Toole, A.L. (1964): S.21 von einem Stichprobenumfang mit $n > 32$. In der Wirtschafts- und Sozialstatistik sind Stichprobenumfänge dieser Größe meist als gegeben zu betrachten.

6.4.2 Erwartungswert bei nicht besetzten Schichten

Dieser Fall, nämlich das Auftreten von sogenannten 0-Schichten, ist bei der A-posteriori-Schichtung kein Non-response-Problem, ganz im Gegensatz zur A-priori-Schichtung (geschichtetes Auswahlverfahren). Als Non-response-Fälle werden diejenigen angesehen, bei denen Einheiten der ausgewählten Stichprobe nicht gemessen (gezählt) werden können. Kurz charakterisiert, sind dies folgende Fälle[1]:

Nichtantreffen der Befragten
Verweigerung der Antwort
Unfähigkeit zur Antwort
Nichtaufsuchen des Befragenden
Verlust von Verzeichnissen

Natürlich können diese Fälle des "Non-response" auch bei der A-posteriori-Schichtung zusätzlich auftreten. Dieses Problem wollen wir hier nicht betrachten, und der Lösungsansatz ist auch ein anderer als der hier behandelte, denn er behandelt unvollständige Stichproben[2]. Wir haben es hier jedoch mit einer vollständigen Stichprobe zu tun, deren allgemeinen Lösungsansatz wir behandeln wollen.

Der nun behandelte Erwartungswert enthält den im letzten Abschnitt behandelten als Spezialfall und/oder gute Approximation. Wie wir nachher sehen werden, interessiert der jetzt behandelte Fall hauptsächlich aus theoretischer Sicht und der Vollständigkeit halber. In der Praxis hat der Statistiker bei der Hochrechnung freie Konzeptwahl. Das bedeutet, daß er von vornherein - wenn er es übersieht und die nötigen Gewichte als zusätzliche Information besitzt - eine entsprechende Schichtung der Daten vornimmt, bzw. noch einmal neu schichtet, da dies allgemein bei der A-posteriori-Schichtung keinen großen Aufwand erfordert. Jedoch wäre diese Lösung ohne den allgemeinen Fall für den Theoretiker unbefriedigend.

Der Lösungsansatz des allgemeinen Falles kann wie folgt gemacht

---

1) Vgl. z.B. Cochran, William G. (1964): S.355ff. und Kish, Leslie (1965): S.362ff..

2) Vgl. Sukhatme, Pandurang V. (1965): S.478ff., ebenso Cochran, William G. (1964): S.367ff..

werden[1]: Werden ein oder mehrere $n_h=0$, so ersetzen wir den fehlenden Schichtmittelwert durch eine willkürlich gewählte Konstante A. Der Einfachheit halber wollen wir für alle h Schichten $A_h=A$ setzen. Diese Vereinfachung beeinflußt die theoretischen Ausführungen nicht. Wir nehmen nun folgende Schätzfunktion an:

$$\bar{y}_p' = \sum_h \frac{N_h}{N} \bar{y}_h'$$

wobei $\bar{y}_h' = \bar{y}_h$, wenn $n_h > 0$

und $\bar{y}_h' = A_h$ bzw. $\bar{y}_h' = A$, wenn $n_h = 0$

Unter der Annahme, daß n_h einer Binomialverteilung der Ordnung n mit der Grundwahrscheinlichkeit $P_h = N_h/N$ und der Gegenwahrscheinlichkeit $Q_h = 1-N_h/N$ genügt, ist bei einem Stichprobenumfang von n die Chance $P(n_h)$ von n_h "Erfolgen". d.h., daß $n_h > 0$ in einer Reihe von n Versuchen ist:

$$P(n_h) = \binom{n}{n_h} P_h^{n_h} \cdot Q_h^{n-n_h}$$

Die mathematische Erwartung des Schichtmittelwerts $\bar{y}_h'$ wird dann als Näherungswert[2]:

$$E(\bar{y}_h') = Q_h^n A + P_h^n \bar{Y}_h = Q_h^n A + (1-Q_h^n)\bar{Y}_h$$

$$= Q_h^n A + \bar{Y}_h - \bar{Y}_h Q_h^n$$

$$= \bar{Y}_h + Q_h^n (A - \bar{Y}_h)$$

Somit wird der Erwartungswert von $\bar{y}_p'$:

$$E(\bar{y}_p') = E\left[\sum_h \frac{N_h}{N} \bar{y}_h'\right] = \sum_h \frac{N_h}{N} E(\bar{y}_h')$$

---

1) Vgl. hierzu Williams, W.H. (1962): S.624.

2) Vgl. Stephan, Frederick F. (1945): S.51/52.

$$= \sum_h \frac{N_h}{N} \bar{Y}_h - \sum_h \frac{N_h}{N} Q_h^n (A-\bar{Y}_h)$$

$$= \bar{Y} + \sum_h \frac{N_h}{N} Q_h^n (A-\bar{Y}_h) \qquad (3)$$

bzw. bei Konstanten A_h für jede Schicht:

$$E(\bar{y}_p') = \bar{Y} + \sum_h \frac{N_h}{N} Q_h^n (A_h-\bar{Y}_h) \qquad (3a)$$

Das Ergebnis zeigt, daß dieser auftretende Bias (vgl. Abschnitt 2.4.5 Bias und Standardabweichung) sehr klein und vom Stichprobenumfang abhängig ist und damit bei größer werdenden n gegen 0 geht; denn die Fälle, bei denen $n_h < 0$, nehmen mit zunehmendem Stichprobenumfang zu. Die allgemeine Schätzfunktion der A-posteriori-Schichtung ist asymptotisch unverzerrt; wir haben es mit einem "asymptotisch biasfreien" Verfahren zu tun.[1] Die beste Wahl für A_h bzw. für A ist der beste Schätzwert für $\bar{Y}_h$, der den Bias gegen 0 gehen läßt.

Setzen wir für $Q_h^n = 1 - \frac{N_h}{N}$, erhalten wir:

$$E(\bar{y}_p') = \bar{Y} + \sum_h \frac{N_h}{N} (1 - \frac{N_h}{N})(A-\bar{Y}_h)$$

$$= \bar{Y} + \sum_h \frac{N_h(N-N_h)}{N^2} (A-\bar{Y}_h)$$

Dieses Ergebnis zeigt sehr deutlich, daß der Bias von sehr kleiner Ordnung ist, da N^2 im Nenner in der Wirtschafts- und Sozialstatistik meist eine große Zahl ist. Als Schätzfunktion erhalten wir also im allgemeinen Fall:

$$\bar{y}_p' = \sum_h \frac{N_h}{N} \bar{y}_h + \sum_h \frac{N_h(N-N_h)}{N^2} (A-\bar{y}_h) \qquad (4)$$

bzw. bei den gewählten Konstanten für jede Schicht:

$$\bar{y}_p' = \sum_h \frac{N_h}{N} \bar{y}_h + \sum_h \frac{N_h(N-N_h)}{N^2} (A_h-\bar{y}_h) \qquad (4a)$$

---

1) Vgl. Abschnitt 2.1.2.3.1 Verzerrte und unverzerrte Schätzung.

Haben wir für jede Schicht den Mittelwert $\bar{y}_h$, d.h., alle n_h sind größer als 0, bzw. haben wir für A_h eine gute Wahl getroffen, so erhalten wir die im letzten Abschnitt behandelte und allgemein übliche (Näherungs-) Schätzfunktion:

$$\bar{y}_p = \sum_h^L \frac{N_h}{N} \bar{y}_h \qquad (1)$$

da der zweite Summenausdruck dann gleich 0 wird.

Neben diesem recht praktikablen Weg zur Lösung des 0-Schichtenproblems kann man andere Lösungswege gehen. So schlägt Fuller[1] vor, die Stichprobe zunächst in zwei Schichten aufzuteilen, diese dann in zwei Unterschichten aufzuteilen und so fort. Daraufhin sucht er, beginnend mit der letzten Aufteilung, für jedes Schichtpaar einen unverzerrten Schätzwert. Hierfür bringt er dann einige umfangreiche Herleitungen. Diese Lösung scheint jedoch nicht günstig zu sein wegen ihres Ansatzes, da die Schichtbildung normalerweise nicht durch weitere Unterteilung der Schichten vorgenommen wird.

3.4.3 Varianz

Bei der Herleitung der Varianz gehen wir von der Formel für die A-priori-Schichtung (geschichtetes Auswahlverfahren) aus[2]:

$$V(\bar{y}_s) = \sum_h^L \frac{N_h^2}{N^2} \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h} \qquad (5)$$

Dabei ist

$$S_h^2 = \frac{\sum_i^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}$$

die wahre Varianz der h-ten Schicht und L die Schichtanzahl.

Der Schätzwert für diese Varianz ist

---

[1]) Fuller, Wayne A. (1966): S.1172ff., insbesondere S.1180ff..

[2]) Die Herleitung der Varianz der A-priori-Schichtung ist in jedem größeren Lehrbuch über Stichproben zu finden.

$$v(\bar{y}_s) = \sum_h \frac{N_h^2}{N^2} \frac{N_h - n_h}{N_h} \frac{s_h^2}{n_h}$$

wobei $$s_h^2 = \frac{\sum_i^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

und $$E(s_h^2) = S_h^2 \quad \text{ist.}$$

Im Gegensatz zum geschichteten Auswahlverfahren (A-priori-Schichtung) ist bei der A-posteriori-Schichtung die Aufteilung der Stichprobeneinheiten auf die einzelnen Schichten nicht bekannt. n_h ist also eine Zufallsvariable. Die Varianz des Schätzwertes $\bar{y}_p$ bzw. $\hat{Y}_p$ der A-posteriori-Schichtung ist eine lineare Funktion der Variablen $1/n_h$. Zur Herleitung formen wir Formel (5) um in

$$V(\bar{y}_s) = \sum_h \frac{N_h^2}{N^2} \frac{S_h^2}{n_h} - \sum_h \frac{N_h^2}{N^2} \frac{S_h^2}{N_h} \qquad (5a)$$

Um zur Varianz der A-posteriori-Schichtung zu gelangen, nehmen wir den Erwartungswert des obigen Ausdrucks (5a), also

$$E\left[V(\bar{y}_s)\right] = V_{(\bar{y}_p)} = E\left[\sum_h \frac{N_h^2}{N^2} \frac{S_h^2}{n_h} - \sum_h \frac{N_h^2}{N^2} \frac{S_h^2}{N_h}\right]$$

$$= \sum_h \frac{N_h^2}{N^2} S_h^2 \, E\left(\frac{1}{n_h}\right) - \sum_h \frac{N_h^2}{N^2} \frac{S_h^2}{N_h} \qquad (6)$$

Wir suchen also den Erwartungswert von $1/n_h$. Zunächst wollen wir ihn in der allgemeinen Form von $1/z$ behandeln.

Die Diskussion um die Berechnung des Erwartungswerts eines Quotienten begann schon Ende des 19.Jahrhunderts, z.B. hat sich K. Pearson 1897 damit befaßt, und dauert bis heute unvermindert an.[1] Nur in Ausnahmefällen ist es möglich die mathematische

---

[1] Tschuprow, A.A. (1925): S.148, der sich ebenfalls damit befaßt hat, bis zu einer umfangreichen Diskussion in The American Statistician, Washington, 1966, Vol.20, Nr.1, S.25; Nr.4, S.28; Nr.5, S.30 und 1967, Vol.21, Nr.2, S.24; Nr.4, S.33.

Erwartung eines Quotienten genau zu berechnen[1]. Im Normalfall geht es darum, gute Approximationen zu erhalten, Anderson[2] geht davon aus, daß

$$E(\frac{1}{z}) = \frac{1}{M_h(z)}$$

d.h., der Erwartungswert von 1/z ist gleich dem Kehrwert des harmonischen Mittels von z. Dieser Ansatz scheint in der Stichprobentheorie nicht allzu fruchtbar, da hier die Schwierigkeiten beginnen, sobald man einen Varianzausdruck berechnen will.

Eine nur schwer verständliche Herleitung einer reziproken Bernoullischen Variablen gibt Stephan[3], jedoch scheint auch dieser Ansatz nicht der geeignete für die Stichprobentheorie zu sein. Ohne jede Herleitung gibt Deming[4] als Näherungswert an

$$E(\frac{1}{z}) \approx \frac{1}{Ez} \left(1 - \frac{\sigma^2 z}{(Ez)^2}\right)$$

Ebenfalls ohne Herleitung geben Cochran, Desabie und Hansen/Hurwitz/Madow ähnliche Ausdrücke für E(1/z) an.[5]

Ein guter Ansatz scheint folgender zu sein[6]:

Es sei $z = Ez + \Delta$

bzw. $\Delta = z - Ez$

mit $E\Delta = 0$

Dann ist

$$\frac{1}{z} = \frac{1}{Ez+\Delta}$$

---

1) Vgl. hierzu Tschuprow, A.A. (1925): S.81ff..

2) Anderson, Oskar N. (1935): S.180ff..

3) Stephan, Frederick F. (1945): S.50-61.

4) Deming, Edwards W. (1960): S.332 und 372.

5) Cochran, Williams G. (1964): S.136; Desabie, J. (1966): S.180; Hansen, Morris H./Hurwitz, William N. /Madow, William G. (1956b): S.139.

6) Vgl. Chiang, C.L. (1966): S.28; ähnlich auch Sukhatme, Pandurang V. (1963): S.51ff..

Aus der Identität[1]

$$\frac{1-a^{n+1}}{1-a} = 1+a+\ldots+a^n \qquad (7)$$

bzw.

$$\frac{1}{1-a} = 1+a+\ldots+a^n + \frac{a^{n+1}}{1-a} \qquad (7a)$$

folgt, wenn wir für

$$a = -\frac{\Delta}{Ez}$$

setzen, die Reihe, die wir bei n=3 abbrechen wollen:

$$\frac{1}{z} = \frac{1}{Ez+\Delta} = \frac{1}{Ez}\left[1 - \frac{\Delta}{Ez} + \frac{\Delta^2}{(Ez)^2} - \frac{\Delta^3}{(Ez)^3} + \frac{\Delta^4}{(Ez)^3(Ez+\Delta)}\right]$$

Der Erwartungswert ist dann:

$$E\left(\frac{1}{z}\right) = \frac{1}{Ez}\, E\left[1 - \frac{\Delta}{Ez} + \frac{\Delta^2}{(Ez)^2} - \frac{\Delta^3}{(Ez)^3} + \frac{\Delta^4}{(Ez)^3(Ez+\Delta)}\right] \qquad (8)$$

Wir sehen sofort, daß die Erwartungswerte der Variablen Δ, wie schon aus der Definition ersichtlich, die Momente einer Verteilung darstellen. Dabei sind das 3. und 4.Moment Maße für Schiefe und Exzeß[2]. Das 3.Moment wird bei einer halbwegs symmetrischen Verteilung gleich 0[3], ebenfalls das 4.Moment, wenn wir es angenähert mit einer Normalverteilung zu tun haben. Bei Gültigkeit des zentralen Grenzwertsatzes und genügend großem Stichprobenumfang können wir diese Voraussetzung machen.

Wir können als Näherungsformel des Erwartungswerts (8) also angeben, da das 1.Moment ($E\Delta$) laut unserer Definition gleich 0 ist:

$$E\left(\frac{1}{z}\right) \approx \frac{1}{Ez}\left(1 + \frac{E\Delta^2}{(Ez)^2}\right) \qquad (9)$$

Nach dieser allgemeinen Behandlung wollen wir zu unserem speziellen Fall der Variablen n_h übergehen. Es ist:

---

1) S. Whitehead, A.N. (1958): S.118ff..

2) Vgl. Smirnow, N.W./Dunin-Barkowski, J.W. (1963): S.73ff..

3) Diese Voraussetzung macht wohl - jedoch stillschweigend - Sukhatme, Pandurang V. (1963): S.107.

$$z = n_h$$

und $$E(n_h) = \frac{n}{N} \cdot N_h \qquad (10)$$

wie sich leicht nachweisen läßt[1].

Ferner ist dann

$$\Delta = n_h - E(n_h) = n_h - \frac{n}{N} \cdot N_h$$

Folglich können wir $E\Delta^2$ als einen Varianzausdruck der Form betrachten

$$E\Delta^2 = \frac{N-n}{N} \cdot n \cdot P(1-P)$$

mit $$P = \frac{N_h}{N}$$

Also $$E\Delta^2 = \frac{N-n}{N} \cdot n \cdot \frac{N_h}{N} \left(1 - \frac{N_h}{N}\right) \qquad (11)$$

Der Ausdruck in der Klammer in Formel (9) wird dann [s. Formel (10) und (11)]:

$$\frac{E\Delta^2}{(En_h)^2} = \frac{1}{(\frac{n}{N} \cdot N_h)^2} \left[\frac{n(N-n)}{N} \frac{N_h}{N} \frac{N-N_h}{N}\right]$$

$$= \frac{N-n}{N \cdot n} \frac{N-N_h}{N_h} \qquad (12)$$

Der gesuchte Erwartungswert von $1/n_h$ wird dann angenähert, indem (10) und (12) in (9) eingesetzt werden:

$$E\left(\frac{1}{n_h}\right) \approx \frac{N}{n \cdot N_h} \left[1 + \frac{N-n}{N \cdot n} \frac{N-N_h}{N_h}\right] \qquad (13)$$

Setzen wir (13) in (6) ein, erhalten wir:

$$V(\bar{y}_p) = \sum_h \frac{N_h^2}{N^2} \frac{N}{nN_h} S_h^2 + \sum_h \frac{N_h^2}{N^2} \frac{(N-n)}{n^2 N_h} \frac{N-N_h}{N_h} S_h^2 - \sum_h \frac{N_h^2}{N^2} \frac{S_h^2}{N_h}$$

---

[1] S. Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956b): S.107.

$$= \frac{1}{n} \sum_h \frac{N_h}{N} S_h^2 - \frac{1}{N} \sum_h \frac{N_h}{N} S_h^2 + \frac{N-n}{Nn^2} \sum_h \frac{N-N_h}{N} S_h^2$$

Nach einigen Umformungen wird dann der endgültige Ausdruck für die Varianz der A-posteriori-Schichtung[1]:

$$V(\bar{y}_p) = \sigma^2(\bar{y}_p) = \frac{N-n}{N \cdot n} \left[\sum_h \frac{N_h}{N} S_h^2 + \frac{1}{n} \sum_h \frac{N-N_h}{N} S_h^2 \right] \qquad (14)$$

Der 1.Summand in der Klammer entspricht der Varianz der Schätzfunktion eines geschichteten Auswahlverfahrens (A-priori-Schichtung) mit proportionaler Aufteilung auf die Schichten. Den 2.Summanden können wir als ein Korrekturglied auffassen, ähnlich jenem bei der Varianz der gruppierten Verfahren (s. Abschnitt 3.3.3 A-posteriori-Schichtung und Gruppierungsverfahren). Daraus zeigt sich, daß die Varianz der A-posteriori-Schichtung tendenziell größer ist als die des geschichteten Auswahlverfahrens; aber der 2.Summand in der Klammer verkleinert sich rasch bei zunehmendem Stichprobenumfang mit $1/n^2$, und die Varianz nähert sich jener bei proportional-geschichtetem Auswahlverfahren (A-priori-Schichtung). Man kann diesen Summanden unter bestimmten Bedingungen vernachlässigen, doch gilt hier ganz besonders der Satz: "but one must be careful"[2]. Darüber

---

[1] Der gleiche Ausdruck ist auch zu finden bei Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956b): S.139 und bei Deming, Edwards W. (1960): S.323, wenn wir $N_h/N=P$ und $(N-N_h)/N=Q$ setzen. Dagegen lassen Cochran, William G. (1964): S.136, Desabie, J. (1966): S.180 und Sukhatme, Pandurang V. (1963): S.107 den Korrekturfaktor (N-n)/N beim 2.Summanden in der Klammer ohne Angabe von Gründen weg, wohl weil die Varianz schon ein Näherungswert ist. Einen wesentlich stärker abweichenden Ausdruck gibt Kish, Leslie (1965): S.90 ohne jegliche Herleitung gleich als Näherungswert an (wobei nicht gesagt wird, wo die Näherung beginnt) mit $\frac{N-n}{N \cdot n} \left[\Sigma \frac{N_h}{N} S_h^2 + \Sigma \frac{N_h}{N} \left(1 - \frac{N_h}{N}\right) \frac{S_h^2}{n_h} \right]$ (Die Notation wurde dabei in die oben gebrauchte gebracht, S.H.) Unserer Ansicht nach gelangt Kish zu dieser Formel mit Hilfe von Analogschlüssen über die Formel für ein proportional-geschichtetes Auswahlverfahren.

[2] Deming, Edwards W. (1960): S.323.

hinaus bedeutet er keine ernste Vergrößerung der Varianz[1], denn wir können - unter Vernachlässigung des Korrekturfaktors für endliche Grundgesamtheiten (N-n)/N - diesen Ausdruck auch umformen in:

$$\frac{1}{n^2} \Sigma \frac{N-N_h}{N} S_h^2 = \frac{1}{n^2} \Sigma S_h^2 - \frac{1}{n^2} \Sigma \frac{N_h}{N} \bar{S}_h^2$$

$$= \frac{1}{n}\left(\frac{L}{n}\right) \bar{S}^2 - \frac{1}{n^2} \Sigma W_h S_h^2$$

$$= \frac{1}{n\bar{n}} \bar{S}^2 - \frac{1}{n^2} \Sigma W_h S_h^2 \qquad (15)$$

Dabei ist $\bar{S}^2 = \Sigma {S_h}^2/L$ die durchschnittliche Varianz der Schichten, $\bar{n} = n/L$ die durchschnittliche Schichtbesetzung und $W_h = N_h/N$. Daraus ist ersichtlich, daß bei nicht allzu differierenden S_h^2 der Ausdruck (15) angenähert gleich 0 wird.

Wir haben es also mit einem angenähert so wirksamen Verfahren zu tun wie beim geschichteten Auswahlverfahren (A-priori-Schichtung), das aber demgegenüber einige Vorteile aufweist (s. Abschnitt 3.2 Die A-posteriori-Schichtung, ein Post-hoc-Verfahren und seine Vorteile).

Die Varianz für den Schätzwert des gesamten Merkmalbetrags $\hat{Y}_p = N\bar{y}_p$ ist dann:

$$V(\hat{Y}_p) = \sigma^2_{(\hat{Y}_p)} = \frac{N-n}{n}\left[\sum_h^L N_h S_h^2 + \frac{1}{n} \sum_h^L (N-N_h) S_h^2\right] \qquad (16)$$

3.4.4 Schätzwert der Varianz

Den Schätzwert der Varianz erhalten wir auf die in der Stichprobenliteratur gehandhabte übliche Weise.[2] Als Schätzwert der

---

[1] Vgl. Cochran, William G. (1964): S.136.

[2] Wir verweisen insbesondere auf die Kapitel über geschichtete Auswahlverfahren, wo auch die Herleitung des unverzerrten Schätzwertes s_h^2 von S_h^2 zu finden ist.

Varianz des Mittelwertes $\bar{y}_p$ bzw. des Totalwertes $\hat{Y}_p$ ergibt sich dann näherungsweise:

$$v(\bar{y}_p) = s^2_{(\bar{y}_p)} = \frac{N-n}{N \cdot n}\left[\sum_h \frac{N_h}{N} s_h^2 + \frac{1}{n}\sum_h \frac{N-N_h}{N} s_h^2\right] \qquad (17)$$

bzw.

$$v(\hat{Y}_p) = \frac{N-n}{n}\left[\sum_h N_h s_h^2 + \frac{1}{n}\sum_h (N-N_h)\, s_h^2\right] \qquad (18)$$

da

$$s_h^2 = \frac{\sum_i^{n_h} (y_{hi}-\bar{y}_h)^2}{n_h-1}$$

der unverzerrte Schätzwert von S_h^2 ist.

Die Formeln für die Vertrauensgrenzen lauten dann wie folgt:
Für den Mittelwert $\bar{Y}$ der Grundgesamtheit:

$$\bar{y}_p \pm t\sqrt{v(\bar{y}_p)}$$

Für den Totalwert Y der Grundgesamtheit:

$$\hat{Y}_p \pm t\sqrt{v(\hat{Y}_p)}$$

Diese Formeln setzen voraus, daß $\bar{y}_p$ - und somit auch $\hat{Y}_p = N\bar{y}_p$ - normalverteilt und $\sqrt{v(\bar{y}_p)}$ hinreichend determiniert ist, so daß der Multiplikator t von den Tabellen der Normalverteilung abgelesen werden kann.

Die Berechnung eines Mindeststichprobenumfanges für einen gegebenen Fehler bei einem vorgegebenen t ist hier in allgemeiner Form sehr umständlich, da wir es mit einer quadratischen Gleichung zu tun haben, deren Koeffizienten aus unhandlichen Ausdrücken bestehen. Im praktischen Fall, wenn Zahlen vorliegen oder beim Näherungswert, wenn der 2.Klammerausdruck weggelassen wird (wie A-priori-Schichtung), läßt sich der Ausdruck einfach lösen.

3.4.5 Bias und Standardabweichung

Mit dem Bias steht und fällt die praktische Bedeutung einer Schätzmethode. Nur wenn plausibel erklärt werden kann, daß der Bias gegenüber der Standardabweichung gering ist, hat der Einsatz wirksamerer (jedoch verzerrter) Schätzverfahren einen Sinn.

Ein Bias tritt auf, wenn

- veraltete Gewichte,
- falsche Gewichte oder
- bei unbekannten Gewichten falsche Schätzungen

zur Aufteilung der Schichten genommen werden[1]. So wurden aus heuristischen Gründen im Rahmen einer Nachprüfung der Angabefehler die Gemeinden in Belgien nach ihrer Schweineanzahl geschichtet[2]. Die Fluktuation dieser Schichten wurde an zwei zeitlich aufeinanderfolgenden Zeitpunkten beobachtet (Landwirtschafts- und Gartenbauzählung vom 15.Dez. 1964 und Landwirtschafts- und Gartenbauzählung vom 15.Mai 1965). Dabei wurde festgestellt, daß sich bei den Gemeinden einer Schicht Mittelwert und Streuung zu den verschiedenen Zeitpunkten wesentlich geändert haben. In 3 von 7 Schichten betrug die Veränderung ungefähr +40%, in 2 Schichten zwischen 10 und 20%, und in der Schicht (Gemeinden mit 100 bis 399 Schweinen) der stärksten Besetzung (annähernd 40% der Gemeinden Belgiens) betrug die Veränderung etwas über +90%. Das bedeutet, daß bei einem geschichteten Auswahlverfahren (A-priori-Schichtung), bei dem die Schichtenbildung am Material der vorhergegangenen Zählung vorgenommen werden muß, ein Bias von erheblicher Größe entstehen kann. Eine optimale Schichtung, die an altem Material vorgenommen worden ist, ist bei strukturellen Veränderungen keine optimale Auswahl für die Auswahl selber. Die Schichtenbildung an altem Material kommt oft vor in der praktischen Statistik, da dies in vielen Fällen die einzige Möglichkeit ist (Wie soll

---

[1] Das gleiche Problem tritt natürlich auch beim geschichteten Auswahlverfahren auf. Auch darüber ist in der Literatur kaum etwas zu finden.

[2] Solche Berechnungen wurden durchgeführt von Herrn Louis Nguiffo, Institut für Wirtschafts- und Sozialstatistik, Tübingen.

man eine Zählung normalerweise anders vorbereiten?). So blieb auch bei der besagten stichprobenweisen Nachprüfung der Angabefehler bei Viehzählungen in Belgien als einzige Möglichkeit, die Schichtung an altem Material vorzunehmen, da die Nachprüfung ungefähr 14 Tage nach der nachzuprüfenden Zählung stattfand. Bei einer A-posteriori-Schichtung wäre es möglich gewesen, die Schichtung nachträglich an Hand der nachzuprüfenden Vollzählung vorzunehmen, somit wäre eine strukturelle Verschiebung der für die Schichtung relevanten Daten ausgeschaltet gewesen.

Wir sehen also, daß wegen der zeitlichen Variabilität der Merkmale eine Schichtung zeitlich so nah wie möglich am Erhebungszeitpunkt stattfinden soll. Das Problem eines möglicherweise auftretenden Bias betrifft hier wohl eher das geschichtete Auswahlverfahren (A-priori-Schichtung) als das geschichtete Hochrechnungsverfahren (A-posteriori-Schichtung). Anders dagegen, wenn zur Schichtbildung falsche Gewichte bzw. Hypothesen genommen werden. Hier tritt das Bias-Problem bei beiden Verfahren auf. Mit der Schichtbildung unter der Annahme bestimmter Hypothesen wollen wir uns jedoch im Kapitel (3.5) beschäftigen.

Den Bias behandeln wir zunächst getrennt von der Varianz. Dies ist zweckmäßig wegen der algebraischen Form der Gleichungen, die den Vergleich der Varianzen und den der Bias' verschiedener Schätzmethoden getrennt zuläßt[1]. Der Bias wird folgendermaßen definiert

$$B(\bar{y}_p) = E\left[\bar{y}_{p_o} - E(\bar{y}_p)\right] \qquad (19)$$

dabei ist

$$E(\bar{y}_p) = \sum_h \frac{N_h}{N} \bar{Y}_h \quad \text{der wahre Wert, und}$$

$$\bar{y}_{p_o} = \sum_h \frac{N_{h_o}}{N_o} \bar{y}_h \quad \text{der Schätzwert, bei dem}$$

die Gewichte N_{h_o}/N_o nicht den wahren Gewichten N_h/N entsprechen. Also:

$$B(\bar{y}_p) = E\left[\sum_h \frac{N_{h_o}}{N_o} \bar{y}_h - \sum_h \frac{N_h}{N} \bar{Y}_h\right]$$

---

[1]) Vgl. Chernoff, Herman (1960): S.154.

$$= \sum_h \frac{N_{h_o}}{N_o} \bar{Y}_h - \sum_h \frac{N_h}{N} \bar{Y}_h$$

$$= \sum_h \left(\frac{N_{h_o}}{N_o} - \frac{N_h}{N}\right) \bar{Y}_h \qquad (20)$$

Es handelt sich hier also nicht um einen Bias, der dem Schätzverfahren innewohnt, sondern er rührt von fehlerhaften Gewichten her.

Ist die Gesamtzahl N der Elemente zwar bekannt, jedoch nicht der Anteil der Schichten, ändert sich (20) in:

$$B'(\bar{y}_p) = \sum_h \frac{N_{h_o} - N_h}{N} \bar{Y}_h \qquad (20a)$$

Als Schätzwert für den Bias nehmen wir:

$$\hat{B}(\bar{y}_p) = \sum_h \left(\frac{N_{h_o}}{N_o} - \frac{N_h}{N}\right) \bar{y}_h \qquad (21)$$

Analog gilt das für Formel (20a).

Der Bias fällt vor allem dann stark ins Gewicht, wenn in allen Schichten eine mehr oder minder gleichmäßige Über- bzw. Unterschätzung der Gewichte erfolgt. Treten jedoch abwechselnd Über- und Unterschätzungen auf, werden sich die einzelnen Schichtwerte des Bias (möglicherweise ganz) ausgleichen. Das ist dann möglich, wenn eine Schicht infolge der fehlerhaften Gewichte überproportional in die Schätzung einbezogen wird, während eine andere nur unterproportional berücksichtigt wird.

Zur Abschätzung der Genauigkeit einer Schätzmethode ist es nötig, den Bias in den Ausdruck für die Varianz einzubauen. Dabei erhalten wir unter Benutzung der Formeln (14) und (20) die Varianzformel für eine verzerrte Schätzung:

$$V'(\bar{y}_p) = \sigma'^2_{(\bar{y}_p)} = \sigma^2_{(\bar{y}_p)} + B^2_{(\bar{y}_p)} \qquad (22)$$

$$= \frac{N-n}{N \cdot n} \left[\sum_h \frac{N_h}{N} s_h^2 + \frac{1}{n} \sum_h \frac{N-N_h}{N} s_h^2 \right] + \left[\sum_h \left(\frac{N_{h_o}}{N_o} - \frac{N_h}{N}\right) \bar{Y}_h \right]^2$$

Der Schätzwert dieser Formel [s. Formeln (17) und (21)] wird:

$$v'(\bar{y}_p) = s^2_{(\bar{y}_p)} + \hat{B}^2_{(\bar{y}_p)} \qquad (23)$$

$$= \frac{N-n}{N\,n}\left[\sum_h^L \frac{N_h}{N}\, s_h^2 + \frac{1}{n}\sum_h^L \frac{N-N_h}{N}\, s_h^2\right] + \left[\sum_h^L \left(\frac{N_{h_o}}{N_o} - \frac{N_h}{N}\right) \bar{y}_h\right]^2.$$

Für die Standardabweichung gilt dann wie üblich

$$\sigma'_{(\bar{y}_p)} = \sqrt{\sigma^2_{(\bar{y}_p)} + B^2_{(\bar{y}_p)}} \qquad (22a)$$

Der Nachweis für die Richtigkeit der Formel (22) kann in folgender Weise erbracht werden[1], wobei wir zur Vereinfachung folgende Annahmen machen:

$$\sigma^2_{(\bar{y}_p)} = E(u - U)^2$$

$$B^2_{(\bar{y}_p)} = E(c - U)$$

Dabei ist U=Eu der Erwartungswert der unverzerrten Schätzung und c der Erwartungswert der nicht erwartungstreuen Schätzfunktion, wobei gemäß unserer Definition der Bias gleich 0 wird bei c=U und somit ein unverzerrter Schätzwert vorliegt. Für die Varianz der verzerrten Schätzung gilt:

$$\sigma'^2 = E(u - c)^2 = \sigma^2_{(u)} + (c-U)^2$$

denn $\quad u-c = u-U+U-c = (u-U) - (c-U)$

$$(u-c)^2 = (u-U)^2 - 2(u-U)(c-U) + (c-U)^2$$

Bilden wir die mathematische Erwartung:

$$E(u-c)^2 = E(u-U)^2 - 2E(u-U)(c-U) + E(c-U)^2$$

dann wird $\quad E(u-U)(c-U) = (c-U)E(u-U) = 0$

---

[1] Vgl. Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956b): S.50ff. oder Sukhatme, Pandurang V. (1963): S.109ff..

Also wäre bewiesen, daß

$$\sigma'^2 = E(u-c)^2 = E(u-U)^2 + E(c-U)^2 = \sigma^2_{(u)} + B^2_{(u)}$$

In unserer alten Terminologie erhalten wir also - jetzt als bewiesene Formel - als Ausdruck für die verzerrte Varianz

$$\sigma'^2_{(\bar{y}_p)} = \sigma^2_{(\bar{y}_p)} + B^2_{(\bar{y}_p)} \qquad (22)$$

Ein Rechenbeispiel mit vereinfachenden Annahmen soll die Größenordnung und Bedeutung des Bias aufzeigen.[1] Aus der Varianzformel bei gruppierten Verfahren (Abschnitt 3.3.3) können wir - sozusagen im umgekehrten Vorgang wie dort, wo wir die Varianz der gruppierten Werte um einen Faktor berichtigen mußten - den Schichtungseffekt herleiten[2]. Die Varianzformel in der Notation der geschichteten Verfahren, bei Vernachlässigung des Korrekturfaktors (N-n)/N, wird dann:

$$S^2 = \Sigma \frac{N_h}{N} S_h^2 + \Sigma \frac{N_h}{N} (\bar{Y}_h - \bar{Y})^2 \qquad (24)$$

Dabei ist der 1.Summand die Varianz bei proportional geschichteten Auswahlverfahren. Der 2.Summand drückt den Schichtungseffekt aus.

In dem Beispiel wollen wir die Annahme machen, daß die A-posteriori-Schichtung dem proportional geschichteten Auswahlverfahren annähernd gleichkommt. Im Abschnitt 3.5 Wirksamkeit wird die Begründung dieser Annahme gezeigt. Bei dem nachfolgenden Beispiel sollen zwei Schichten gebildet werden mit den fehlerfreien Gewichten

$$\frac{N_1}{N} = 0{,}9; \quad \frac{N_2}{N} = 0{,}1$$

---

[1] Vgl. Cochran, William G. (1964): S.116ff., der das Beispiel für das geschichtete Auswahlverfahren rechnet.

[2] Eine mehr formale Herleitung ist zu finden bei: Cochran, William G. (1964): S.98f. und Kellerer, Hans (1963): S.113, wobei als Näherung gilt $N_h - 1 = N_h$ und $N - 1 = N$. Die Herleitung erfolgt dort zur Feststellung der Wirksamkeit des geschichteten Auswahlverfahrens.

und den fehlerhaften Gewichten

$$\frac{N_{1o}}{N_o} = 0{,}92; \quad \frac{N_{2o}}{N_o} = 0{,}08$$

Weiter wollen wir ein Verfahren

a) mit geringem,

b) mit großem Schichtungseffekt

durchführen:

a) $\bar{Y}_1 - \bar{Y}_2 = 1$

b) $\bar{Y}_1 - \bar{Y}_2 = 3$

Zur Berechnung nehmen wir folgende Vereinfachungen vor:

$$S^2 = 1$$

$$S_1^2 = S_2^2 = S_h^2$$

Bei zwei Schichten nimmt Formel (24) folgende Gestalt an:

$$S^2 = \frac{N_1 S_h^2 + N_2 S_h^2}{N} + \frac{N_1 N_2}{N^2} (\bar{Y}_1 - \bar{Y}_2)^2$$

Setzen wir die obigen Werte ein, erhalten wir im Fall

a) $S_h^2 = 0{,}91$

d.h. eine Reduzierung der Varianz durch die Schichtung um 9%; im Fall

b) $S_h^2 = 0{,}19$

d.h. eine Reduzierung der Varianz durch die Schichtung um 81%.

Die Formel für den Bias beim Zwei-Schichten-Fall wird [s. Formel (20)]:

$$B = (\frac{N_{1o}}{N_o} - \frac{N_1}{N})\bar{Y}_1 + (\frac{N_{2o}}{N_o} - \frac{N_2}{N})\bar{Y}_2$$

$$= (\frac{N_{1o}}{N_o} - \frac{N_1}{N})\bar{Y}_1 + \left[(1 - \frac{N_{1o}}{N_o}) - (1 - \frac{N_1}{N})\right]\bar{Y}_2$$

$$= (\frac{N_{1o}}{N_o} - \frac{N_1}{N})(\bar{Y}_1 - \bar{Y}_2)$$

Setzen wir die angegebenen Werte in die Formel ein, erhalten

wir für den Bias in

a) $B = 0{,}02$

und

b) $B = 0{,}06$.

Der Ausdruck für Varianz und Bias einer verzerrten A-posteriori-Schichtung (wieder bei Vernachlässigung des Korrekturfaktors $(N-n)/N$) wird dann laut Formel (22) bei unseren vereinfachenden Annahmen:

$$V'(\bar{y}_p) = \frac{1}{n}\left[\frac{N_1}{N} S_h^2 + \frac{N_2}{N} S_h^2\right] + \frac{1}{n^2}\left[(1-\frac{N_1}{N})S_h^2 + (1-\frac{N_2}{N})S_h^2\right]$$

$$+ \left[(\frac{N_{1o}}{N_o} - \frac{N_1}{N})(\bar{Y}_1-\bar{Y}_2)\right]^2$$

$$= S_h^2(\frac{1}{n} + \frac{1}{n^2}) + \left[(\frac{N_{1o}}{N_o} - \frac{N_1}{N})(\bar{Y}_1-\bar{Y}_2)\right]^2$$

Setzen wir die angegebenen Werte ein, liefert uns der Fall

a) $V'(\bar{y}_p) = 0{,}91(\frac{1}{n} + \frac{1}{n^2}) + 0{,}0004$ und Fall

b) $V'(\bar{y}_p) = 0{,}19(\frac{1}{n} + \frac{1}{n^2}) + 0{,}0036$

Bei unseren Annahmen wird die Varianz der einfachen Zufallsauswahl, die wir zum Vergleich heranziehen wollen:

$$V(\bar{y}) = \frac{S^2}{n} = \frac{1}{n}$$

Ein Vergleich der Varianzen bei zunehmendem Stichprobenumfang n ergibt folgende Tabelle[1]:

---

[1]) Vergleicht man diese Tabelle mit jener, die Cochran, William G. (1964): S.117 für das proportional geschichtete Auswahlverfahren aufgestellt hat, so zeigt sich - wie schon früher angemerkt wurde und bei der Wirksamkeitsbetrachtung (Abschnitt 3.5) bestätigt werden wird -, daß die A-posteriori-Schichtung in der Wirksamkeit dem proportional geschichteten Auswahlverfahren gleichkommt.

Tabelle 1 Vergleich der Varianzen der einfachen Hochrechnung mit einer a) weniger, b) sehr wirksamen A-posteriori-Schichtung, die einen Bias aufweist.

Stichprobenumfang n	einfache Hochrechnung $V(\bar{y})$	A-posteriori-Schichtung $V'(\bar{y}_p)$ a) weniger wirksam	b) sehr wirksam
30	0,0333	0,0317	0,0101
50	0,0200	0,0190	0,0075
100	0,0100	0,0096	0,0055
200	0,0050	0,0050	0,0046
300	0,0033	0,0035	0,0042
400	0,0025	0,0027	0,0041
1000	0,0010	0,0013	0,0038

Die Tabelle zeigt, daß in unserem Beispiel die A-posteriori-Schichtung in beiden Fällen a) und b) bis zu n=200 wirksamer ist, obwohl sie einen beträchtlichen Bias aufweist. Bei n=300 macht sich die konstante Größe des Bias bemerkbar, der bei zunehmendem Stichprobenumfang immer deutlicher hervortritt. Vergleichsweise wäre die Varianz bei n=1000 im Falle b) ohne Bias $V(\bar{y}_p) = 0,0002$, d.h. 1/5 des Betrages der Varianz der einfachen Hochrechnung, so aber bewirkt der Bias, daß diese die 4fache Größe von jener der einfachen Hochrechnung annimmt. Unser Beispiel, in dem wir relativ extreme Annahmen unterstellt haben, zeigt, daß die A-posteriori-Schichtung selbst mit einem Bias meist wirksamer ist als die einfache Hochrechnung und daß der konstante Bias erst bei großen Stichprobenumfängen die höhere Wirksamkeit zunichte macht.

3.4.6 Die A-posteriori-Schichtung bei gleich großen Schichten

Es mag vorkommen, daß man in der Praxis gleich große Schichten bilden kann. Darüber hinaus können diese sich vereinfachenden Formeln für eine Überschlagsrechnung sehr nützlich sein. Es gilt nun:

$$\left.\begin{aligned} \frac{N_h}{N} &= p_o = \text{const.} \\ 1 - \frac{N_h}{N} &= q_o = \text{const.} \end{aligned}\right\} \text{"wahre" Anteilssätze}$$

$$\left.\begin{aligned} \frac{N_{h_o}}{N_o} &= p'_o = \text{const.} \\ 1 - \frac{N_{h_o}}{N_o} &= q'_o = \text{const.} \end{aligned}\right\} \text{"falsche" (mit Bias behaftete) Anteilssätze}$$

Die Schätzformel für den Mittelwert wird dann bei gleichen Schichten:

$$\bar{y}_{po} = p_o \sum_h \bar{y}_h \qquad (1a)$$

Der Schätzwert der Varianz wird dann zu:

$$v(\bar{y}_{po}) = \frac{N-n}{N \cdot n} \left(p_o \sum_h s_h^2 + \frac{q_o}{n} \sum_h s_h^2\right) = \frac{N-n}{N \cdot n} \left(p_o + \frac{q_o}{n}\right) \sum_h s_h^2 \qquad (14a)$$

Auch hier zeigt sich wieder, daß der 2.Summand der Varianz mit dem Faktor q_o/n sehr klein wird und in vielen Fällen vernachlässigbar ist.

Der Schätzwert für den Bias vereinfacht sich zu folgender Form:

$$\hat{B}(\bar{y}_{po}) = (p'_o - p_o) \sum_h \bar{y}_h = c \Sigma \bar{y}_h \qquad (20a)$$

wobei $\qquad c = p'_o - p_o$

Also wird der Ausdruck für den Schätzwert der Varianz (14a) bei Berücksichtigung des Bias (20a) und gleich großen Schichten:

$$v'(\bar{y}_{po}) = v(\bar{y}_{po}) + \hat{B}^2(\bar{y}_{po}) \qquad (23a)$$

$$= \frac{N-n}{N \cdot n} \left(p_o + \frac{q_o}{n}\right) \Sigma s_h^2 + c^2 (\Sigma \bar{y}_h)^2$$

3.5 Wirksamkeit

Zum Vergleich der Wirksamkeit betrachten wir die bekanntesten - in Kapitel 2.3 behandelten Hochrechnungsarten - und kurz auch das geschichtete Auswahlverfahren (A-priori-Schichtung). Beim Effizienzvergleich gehen wir von den Varianzformeln aus und vernachlässigen den Bias. Dies ist von der algebraischen Form der Gleichungen her möglich[1]; der tiefere Grund ist jedoch, daß sich über die Größe des Bias theoretisch wenig sagen läßt. Hier kommt es auf den praktischen Einzelfall der vorliegenden Daten an. Vor allem läßt sich bei Berücksichtigung des Bias noch weniger eindeutig eine Rangfolge von auf- bzw. absteigend wirksamen Schätzmethoden aufstellen. Bei Berücksichtigung des Bias müssen wir uns darauf beschränken, einige allgemein abwägende Bemerkungen zu machen und können nicht, wie wir es im folgenden versuchen wollen, eine Rangfolge der Schätzmethoden bezüglich ihrer Wirksamkeit bilden.

Zunächst wollen wir den Aspekt der Wirksamkeit einer theoretischen Betrachtung unterziehen und ihn dann an einem Rechenbeispiel praktisch aufzeigen. Zusätzliche Bemerkungen zur Wirksamkeit sollen dann noch bei der Behandlung der kombinierten Hochrechnungsverfahren in Abschnitt 3.8 gemacht werden. Dabei bilden wir Verfahren aus der Kombination der A-posteriori-Schichtung mit der Verhältnis-, Differenzen- und Regressionsschätzung.

3.5.1 Theoretische Betrachtung

Von zwei Schätzmethoden soll jene wirksamer sein, die die geringere Varianz aufweist. Die Varianzausdrücke der einzelnen Hochrechnungsarten sind in den Kapiteln 2.3 und 3.4.3 dargestellt. Wir vergleichen die A-posteriori-Schichtung einzeln mit den anderen Hochrechnungsverfahren und kurz auch mit dem geschichteten Auswahlverfahren (A-priori-Schichtung) und stellen die Ergebnisse dann in einer Tabelle zusammen.

---

[1]) Vgl. Chernoff, Herman (1960): S.154.

3.5.1.1 Vergleich mit der einfachen Hochrechnung und dem geschichteten Auswahlverfahren

Die Varianz des Mittelwertes der einfachen Hochrechnung ist:

$$V(\bar{y}) = \frac{N-n}{N} \frac{S^2}{n}$$

Bei genügend großen N_h und somit auch N können wir als Vereinfachung annehmen[1]:

$$\frac{N_h-1}{N_h} \approx 1 \qquad \text{und} \qquad \frac{N-1}{N} \approx 1$$

Dann können wir den obigen Varianzausdruck umformen in:

$$V(\bar{y}) = \frac{N-n}{N \cdot n}\left[\Sigma \frac{N_h}{N} S_h^2 + \frac{\Sigma N_h(\bar{Y}_h-\bar{Y})^2}{N}\right] \qquad (24)$$

Zum Vergleich der Wirksamkeit bilden wir die Differenz der Varianzen und erhalten:

$$V(\bar{y}) - V(\bar{y}_p) = \frac{N-n}{N \cdot n}\left[\frac{\Sigma N_h(\bar{Y}_h-\bar{Y})^2}{N} - \frac{1}{n}\Sigma \frac{N-N_h}{N} S_h^2\right]$$

oder

$$V(\bar{y}) = V(\bar{y}_p) + \frac{N-n}{N \cdot n}\left[\frac{\Sigma N_h(\bar{Y}_h-\bar{Y})^2}{N} - \frac{1}{n}\Sigma \frac{N-N_h}{N} S_h^2\right] \qquad (25)$$

Wie wir an früherer Stelle bemerkt haben, geht der 2.Summand in der Klammer mit 1/n rasch gegen 0, das bedeutet, daß schon bei geringstem Schichtungseffekt die A-posteriori-Schichtung wirksamer ist als die einfache Hochrechnung. Je größer dieser ist, d.h. je mehr sich die Mittelwerte $\bar{Y}_h$ voneinander unterscheiden und je homogener die Schichten sind, um so größer ist die Wirksamkeit.

Die Varianz beim geschichteten Auswahlverfahren (A-priori-Schichtung) bei proportionaler Aufteilung ist[2]:

$$V(\bar{y}_s) = \frac{N-n}{N \cdot n} \Sigma \frac{N_h}{N} S_h^2$$

---

[1]) Vgl. z.B. Sukhatme, Pandurang V. (1963): S.92.

[2]) Vgl. z.B. Cochran, William G. (1964): S.91.

Zum Vergleich formen wir die Varianz der A-posteriori-Schichtung um:

$$V(\bar{y}_p) = \frac{N-n}{N \cdot n}\left[\Sigma \frac{N_h}{N} S_h^2 + \frac{1}{n}\Sigma \frac{N-N_h}{N} S_h^2\right] \qquad (14)$$

$$= V(\bar{y}_s) + \frac{N-n}{N} \frac{1}{n^2} \Sigma \frac{N-N_h}{N} S_h^2 \qquad (26)$$

Wie leicht ersichtlich, strebt der zweite Faktor schnell gegen 0, das zeigt, daß beide Verfahren angenähert gleich wirksam sind.

Wir können also folgende Rangordnung der Wirksamkeit aufstellen:

$$V(\bar{y}_s) \approx V(\bar{y}_p) < V(\bar{y})$$

oder die strenge und weniger eindeutige Beziehung

$$V(\bar{y}_s) \lesseqgtr V(\bar{y}_p) \lesseqgtr V(\bar{y})$$

3.5.1.2 Vergleich mit der Verhältnisschätzung

Die Varianz der Verhältnisschätzung ist:

$$V(\bar{y}_v) = \frac{N-n}{N \cdot n}\left[S_y^2 + F^2 S_x^2 - 2F\varrho S_x S_y\right]$$

oder da S_y^2 dem S^2 der einfachen Hochrechnung entspricht, können wir bilden [s. Formel (24)]:

$$V(\bar{y}_v) = \frac{N-n}{N \cdot n}\left[\Sigma \frac{N_h}{N} S_h^2 + \frac{\Sigma N_h (\bar{Y}_h - \bar{Y})^2}{N} + F^2 S_x^2 - 2F\varrho S_x S_y\right]$$

Bilden wir wieder die Differenz der Varianzen, ergibt sich

$$V(\bar{y}_v) - V(\bar{y}_p) = \frac{N-n}{N \cdot n}\left\{\left[\frac{\Sigma N_h (\bar{Y}_h - \bar{Y})^2}{N} + F^2 S_x^2 - 2F\varrho S_x S_y\right] - \frac{1}{n}\Sigma \frac{N-N_h}{N} S_h^2\right\} \qquad (27)$$

Der zweite Ausdruck in der geschweiften Klammer mit 1/n ist von geringer Größe, wie schon öfters ausgeführt. Wenn nun die

$\bar{Y}_h$ sehr verschieden sind (großer Schichtungseffekt), dann ist die A-posteriori-Schichtung wirksamer als die Verhältnisschätzung. Deutlicher sehen wir das, wenn wir folgende Beziehung ansetzen:

$$V(\bar{y}_v) > V(\bar{y}_p)$$

bzw. die Ausdrücke eingesetzt und umgeformt:

$$\frac{\Sigma N_h(\bar{Y}_h-\bar{Y})^2}{N} + F^2 S_x^2 - 2F\varrho S_x S_y > \frac{1}{n}\Sigma \frac{N-N_h}{N} S_h^2 \qquad (27a)$$

Dabei ist der Ausdruck auf der rechten Seite der Ungleichung sehr klein.

Ein Vergleich beider Verfahren ist sehr schwierig, das gilt gleichermaßen für die Differenzen- und Regressionsschätzung, weil die Merkmale der zusätzlichen Information von verschiedener Art sind. Bei der A-posteriori-Schichtung ist das Merkmal der zusätzlichen Information von qualitativer Art, während bei den anderen Verfahren die zusätzliche Information von quantitativen Merkmalen geliefert wird.[1] Grundsätzlich können wir sagen, daß eine höhere Wirksamkeit der A-posteriori-Schichtung ausschließlich von der Größe des Schichtungseffekts abhängt. Die zusätzliche Information, die uns die Gewichte (Schichtanteilssätze) liefert, ist um so wertvoller und führt zu einer größeren Effizienz, je klarer sich die Schichten trennen lassen und je homogener die Merkmale innerhalb der Schichten sind.

3.5.1.3 Vergleich mit der Differenzenschätzung

Da die Differenzenschätzung viele Berührungspunkte mit der Verhältnisschätzung besitzt (s. Abschnitt 2.3.5), gelten die oben gemachten Bemerkungen auch hier. Zur formalen Herleitung des Wirksamkeitsvergleichs brauchen wir die Varianz der Differenzenschätzung:

---

[1]) Die Analogie zwischen Verhältnisschätzung und A-posteriori-Schichtung wird von Thionet, M.P. (1953a): S.246 stark betont, jedoch mit dem oben erwähnten Unterschied im quantitativen bzw. qualitativen Charakter der Merkmale.

$$V(\bar{y}_d) = \frac{N-n}{N \cdot n} (S_y^2 + k^2 S_x^2 - 2k\varrho S_x S_y)$$

Wegen der analogen Form der Varianzformeln von Verhältnis- und Differenzenschätzung können wir sofort schreiben:

$$V(\bar{y}_d) > V(\bar{y}_p)$$

bzw. wieder eingesetzt:

$$\frac{\Sigma N_h (\bar{Y}_h - \bar{Y})^2}{N} + k^2 S_x^2 - 2k\varrho S_x S_y > \frac{1}{n} \Sigma \frac{N-N_h}{N} S_h^2 \qquad (28)$$

Auch hier gilt wieder, daß die A-posteriori-Schichtung nur bei Vorhandensein eines genügend großen Schichtungseffekts wirksamer ist als die Differenzenschätzung.

3.5.1.4 Vergleich mit der Regressionsschätzung

Der Näherungswert der Varianz der Regressionsschätzung ist:

$$V(\bar{y}_r) = \frac{N-n}{N \cdot n} S_y^2 (1-\varrho^2)$$

$$= \frac{N-n}{N \cdot n} (S_y^2 - S_y^2 \varrho^2)$$

Wir formen $S_y^2 = S^2$ wieder um [s. (24)] und bilden die Differenz der Varianzen:

$$V(\bar{y}_r) - V(\bar{y}_p) = \frac{N-n}{N \cdot n} \left\{ \left[\Sigma \frac{N_h}{N} S_h^2 + \frac{\Sigma N_h (\bar{Y}_h - \bar{Y})^2}{N} - S_y^2 \varrho^2 \right] - \left[\Sigma \frac{N_h}{N} S_h^2 + \frac{1}{n} \Sigma \frac{N-N_h}{N} S_h^2 \right] \right\}$$

$$= \frac{N-n}{N \cdot n} \left[\frac{\Sigma N_h (\bar{Y}_h - \bar{Y})^2}{N} - S_y^2 \varrho^2 - \frac{1}{n} \Sigma \frac{N-N_h}{N} S_h^2 \right] \qquad (29)$$

Hier können wir schon sehen, daß bei einigermaßen hohem Korre-

lationskoeffizienten die Regressionsschätzung wirksamer ist als die A-posteriori-Schichtung, da der den Schichtungseffekt darstellende Ausdruck praktisch um die Varianz der einfachen Hochrechnung und um einen weiteren Faktor vermindert wird. Zur Verdeutlichung dieses Tatbestandes nehmen wir noch einige Vereinfachungen vor. Der Korrelationskoeffizient soll angenähert gleich 1 sein, und den Faktor $\frac{1}{n}\Sigma \frac{N-N_h}{N} S_h^2$ wollen wir angenähert gleich 0 setzen. Dabei wirkt die Annahme über den Korrelationskoeffizienten zugunsten der Regressionsschätzung, während die 2.Annahme zugunsten der A-posteriori-Schichtung wirkt. Wir erhalten dann, wenn wir wieder die öfters gebrauchte Beziehung (24) für $S_y^2 = S^2$ einsetzen:

$$V(\bar{y}_r) - V(\bar{y}_p) \approx \frac{N-n}{N \cdot n}\left[\frac{\Sigma N_h(\bar{Y}_h-\bar{Y})^2}{N} - \Sigma \frac{N_h}{N} S_h^2 - \frac{\Sigma N_h(\bar{Y}_h-\bar{Y})^2}{N}\right]$$

Also

$$V(\bar{y}_r) \approx V(\bar{y}_p) - \frac{N-n}{N \cdot n} \Sigma \frac{N_h}{N} S_h^2 \qquad (29a)$$

Die Regressionsschätzung ist also wirksamer als die A-posteriori-Schichtung. Es gilt:

$$V(\bar{y}_r) < V(\bar{y}_p)$$

3.5.1.5 Zusammenfassende Darstellung

Vergleichen wir die 4 Hochrechnungsarten, so erhalten wir das Ergebnis, daß sich die Vernachlässigung eines genügend großen Schichtungseffekts außer bei der Regressionsschätzung in allen Fällen nachteilig auf die Effizienz auswirkt. Dies ist um so bedeutsamer, als der zusätzliche Rechenaufwand bei der A-posteriori-Schichtung gegenüber den anderen gebundenen Hochrechnungsarten minimal ist. Aber, um es noch einmal klar herauszustellen, diese Aussage steht und fällt mit der Größe des Schichtungseffekts. Nur bei genügend großem Schichtungseffekt können wir angenähert folgende Rangfolge bezüglich der Effizienz aufstellen:

$$V(\bar{y}_r) \leqq V(\bar{y}_p) \leqq V(\bar{y}_v) \lessgtr V(\bar{y}_d) \leqq V(\bar{y})$$

Dieser angestellte Wirksamkeits- bzw. Varianzenvergleich ist nicht problemlos. Wir haben zwar an früherer Stelle festgestellt, daß die A-posteriori-Schichtung zu den Hochrechnungsverfahren gezählt wird, und daß ferner ihre Wirksamkeit vom Verteilungstyp abhängig ist. Im Vergleich zu den anderen gebundenen Hochrechnungsverfahren ist die A-posteriori-Schichtung von einer völlig anderen Art. Ihre Wirksamkeit ist abhängig davon, ob eine Schichtung vorliegt oder nicht, d.h., sie ist von qualitativer Art. Bei den anderen gebundenen Hochrechnungsverfahren kann man die Wirksamkeit gegenüber der einfachen Hochrechnung ausdrücken in Abhängigkeit von ϱ, dem Korrelationskoeffizienten zwischen y (untersuchte Variable) und x (Variable der zusätzlichen Information). Hier kann man aber nicht sagen, die Wirksamkeit hänge ab vom Vorliegen eines Korrelationskoeffizienten, sondern sie hängt von seiner Größe ab; sie ist also von quantitativer Art.

Allgemein kann man sagen, die unterschiedliche Wirksamkeit von Regressions-, Verhältnis- und Differenzenschätzung hängt ab von den Koeffizienten b, f und k, wenn man sie nicht, wie oben angedeutet, in Abhängigkeit von ϱ ausdrücken will. Die drei Verfahren haben die gleiche Wirksamkeit, wenn gilt[1]:

$$b = f = k$$

Da b normalerweise der bestangepaßte Koeffizient für die Schätzung ist, gilt die Regressionsschätzung als die wirksamste Schätzmethode, was auch ein Vergleich der Formeln zeigt.

Problematischer ist der Vergleich zwischen Verhältnisschätzung und Differenzenschätzung. Hier läßt sich nicht allgemein sagen: wenn $k \gtrless f$, dann ist auch $V(\bar{y}_v) \gtrless V(\bar{y}_d)$ oder umgekehrt.

Die beiden Varianzausdrücke waren:

$$V(\bar{y}_v) = \frac{N-n}{N \cdot n} (S_y^2 + F^2 S_x^2 - 2FS_{xy})$$

$$V(\bar{y}_d) = \frac{N-n}{N \cdot n} (S_y^2 + k^2 S_x^2 - 2kS_{xy})$$

---

[1]) Vgl. Hauser, Siegfried (1965)(unveröff.): S.132ff. und S.151ff..

Wir sehen, daß beide Varianzausdrücke die gleiche Form haben. Bei Vernachlässigung des Korrekturfaktors für endliche Grundgesamtheiten können wir dafür schreiben:

$$z = \alpha + \beta u^2 - 2\gamma u$$

Dabei ist

$$\alpha = S_y^2$$

$$\beta = S_x^2$$

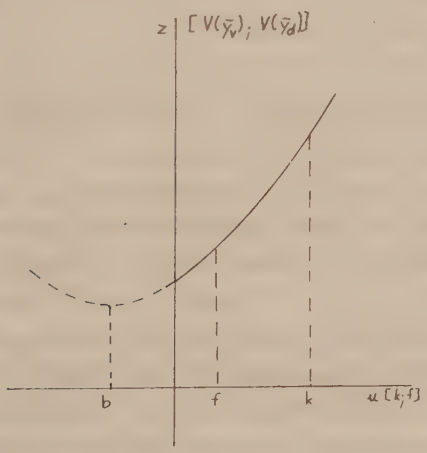
$$\gamma = S_{xy}$$

$$u = k \text{ bzw. } f$$

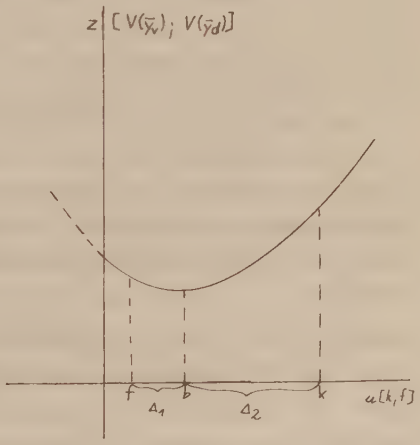
Geometrisch stellt die Funktion die Form einer Parabel dar, deren Tiefpunkt bei

$$u = \frac{\gamma}{\beta} = \frac{S_{xy}}{S_x^2} = b$$

dem Regressionskoeffizienten liegt. Beim Vergleich der Größe der Varianzen in Abhängigkeit von k bzw. f lassen sich zunächst 2 Fälle a) und b) unterscheiden, wobei dann b) mehrere Unterfälle aufweist.



a) Verlauf der Kurve bei negativem b



b) Verlauf der Kurve bei positivem b (Fall b_3)

Im Fall a) liegt der Tiefpunkt auf der negativen Abszisse oder bei Null:

$$b = \frac{S_{xy}}{S_x^2} \leqq 0$$

dann gilt immer, bei k>f ist auch $V(\bar{y}_d) > V(\bar{y}_v)$ und umgekehrt. Dies ist aus dem Schaubild a) leicht zu ersehen.

Im Fall b) liegt der Tiefpunkt auf dem positiven Teil der Abszisse:

$$b > 0$$

Hier lassen sich nun 4 Unterfälle unterscheiden:

b_1) Sind k und f größer als b (beide Punkte liegen auf dem ansteigenden Ast der Parabel), dann gilt das gleiche wie bei a)
bei k>f ist auch $V(\bar{y}_d) > V(\bar{y}_v)$ und umgekehrt.

b_2) Sind k und f kleiner als b (beide Punkte liegen auf dem fallenden Ast der Parabel), dann gilt
bei k>f ist $V(\bar{y}_d) < V(\bar{y}_v)$ und umgekehrt.

b_3) In diesem Fall soll f<b und k>b sein, dann sind die Differenzen $\Delta_1 = b-f$ und $\Delta_2 = k-b$ zu vergleichen (s. Schaubild b). Ist nun
$\Delta_1 < \Delta_2$, dann ist $V(\bar{y}_d) > V(\bar{y}_v)$ und bei
$\Delta_1 > \Delta_2$, dann ist $V(\bar{y}_d) < V(\bar{y}_v)$.

b_4) Das ist der symmetrische Fall zu b_3. Hier gilt Entsprechendes.

Da beim Vergleich der Wirksamkeit bezüglich der Verhältnis- und Differenzenschätzung nur von Fall zu Fall unterschieden werden kann, ist diese ganze Betrachtung als Aussage nicht befriedigend.

Beim Vergleich der gebundenen Hochrechnungsarten mit der einfachen Hochrechnung gilt, daß erstere normalerweise wirksamer sind, da bei ihnen zwei Informationen y und x in der Schätzung verarbeitet werden, und diese somit suffizientere Schätzverfahren sind als die einfache Hochrechnung. Zum Teil haben wir den formalen Vergleich an früherer Stelle schon durchgeführt, für den anderen Teil sei auf die Literatur verwiesen.

Die Stellung der A-posteriori-Schichtung in der Rangfolge der Wirksamkeit der Schätzverfahren hängt von ihrem Schichtungseffekt ab. Bei geringem Schichtungseffekt gilt die obige Rangfolge in geänderter Form:

$$V(\bar{y}_r) \leqq V(\bar{y}_v) \lessgtr V(\bar{y}_d) \leqq V(\bar{y}_p) \leqq V(\bar{y})$$

An Hand des folgenden Beispiels soll die Wirksamkeit der A-posteriori-Schichtung demonstriert werden.

3.5.2 Rechenbeispiel zur Demonstration der Wirksamkeit

Im folgenden Beispiel nehmen wir an, wir hätten in einem Land im Juni 1968 auf Grund der Gartenbauerhebung N=264 Apfelgärten gezählt. Aus dieser Zählung ergab sich die Zahl der Apfelbäume mit X=22 481. Wir wollen nun im Herbst 1968 die Apfelernte des Landes mit Hilfe einer Stichprobe schätzen. Dazu wählen wir n=100 Gärten aus, bei denen wir den Obstertrag messen. Die zusätzliche Information bei den gebundenen Hochrechnungsarten sei der Baumbestand. In dem Beispiel wurde bei den 264 Gärten ein unterschiedlicher Baumbestand angenommen, der bei Obstanlagen, die im Nebenerwerb betrieben werden, aus 24 bis 42 Bäumen je Garten bestand. Diese Gruppe umfaßt mit N_1=200 die Mehrzahl der Betriebe. Daneben wurde N_2=64 große Obstanlagen gezählt, deren Baumbestand zwischen 200 und 290 Bäumen lag. Um es wirklichkeitsgetreu zu gestalten, wurden die 264 Gärten mit ihrem unterschiedlichen Baumbestand in eine Art Gartenbauliste eingetragen. Dann wurden 100 Gärten mit Hilfe des Zufallsprinzips (Lose) ausgewählt. Um die umständliche Arbeit zu vermeiden, allen 22 481 Bäumen einen Ertrag zuzuordnen, wurde nur den Bäumen der ausgewählten Gärten per Zufall ein Obstertrag zwischen 70 und 130kg zugeordnet.

Für die Wirksamkeit der A-posteriori-Schichtung ist die Häufigkeitsverteilung ausschlaggebend. Das folgende Schaubild soll die Stichprobenverteilung aufzeigen, wobei die normierten Häufigkeiten wegen der unterschiedlichen Größe der Zahlen in lo-

garithmischer Form dargestellt werden. Diese halblogarithmische Darstellung läßt die Schichtung nicht so deutlich zum Ausdruck kommen.

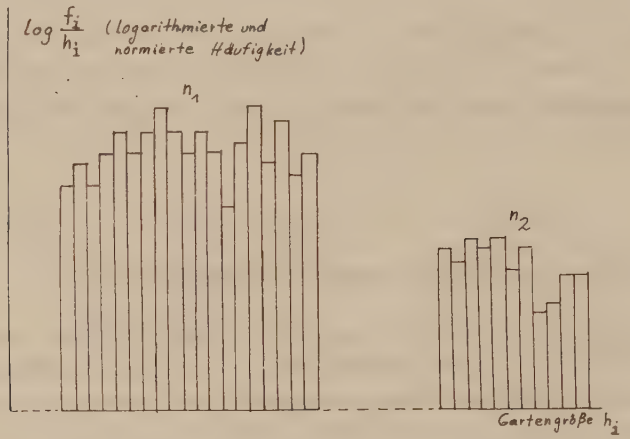


Abb.11 Häufigkeitsverteilung der Stichprobendaten

Ähnliche Verteilungen - meist mit mehreren Schichten - finden wir auch auf anderen Gebieten der Landwirtschaft, bei den Umsätzen der Unternehmen und bei der personellen Einkommensverteilung[1)].

In den mit Hilfe der einfachen Zufallsauswahl bestimmten n=100 Gärten erhielten wir durch Messen und Zählen eine durchschnittliche Ernte je Garten mit

$$\bar{y} = 87{,}77 \text{ dz Äpfel}$$

und einen durchschnittlichen Baumbestand (zusätzliche Information) je Garten von

$$\bar{x} = 89 \text{ Apfelbäumen.}$$

Als Schätzwerte der Varianzen wurden berechnet für die Erträge der einzelnen Gärten in der Stichprobe:

$$s_y^2 = 9174{,}43$$

---

1) Vgl. z.B. Krelle, Wilhelm (1962): S.146; ebenso die Zahlenangaben zu den obengenannten Gebieten im Statistischen Jahrbuch für die Bundesrepublik Deutschland.

für die Anzahl der Bäume:

$$s_x^2 = 8415{,}87$$

Ferner beträgt die Kovarianz:

$$s_{xy} = 8415{,}87$$

Der Bravais-Pearsonsche Korrelationskoeffizient zwischen Ertrag und Bäumen ist:

$$r = 0{,}97$$

Im folgenden betrachten wir nun die Ergebnisse der einzelnen Hochrechnungsarten.

3.5.2.1 Einfache Hochrechnung

Der Schätzwert ist:

$$\hat{Y} = N\bar{y} = 23\ 171 \text{ dz}$$

Der Schätzwert der Varianz beträgt:

$$v(\hat{Y}) = \frac{N(N-n)}{n} s_y^2 = 3\ 972\ 161$$

und die Standardabweichung ergibt dann:

$$s_{\hat{Y}} = 1\ 993 \quad (\text{dz})$$

Also liegt der Wert der Grundgesamtheit mit einer Wahrscheinlichkeit $\Phi(2) = 95{,}45\%$ zwischen den Grenzen

$$19\ 185 \leqq Y \leqq 27\ 157$$

3.5.2.2 A-posteriori-Schichtung

Für die A-posteriori-Schichtung teilen wir die Stichprobendaten - wie aus Abbild 11 ersichtlich - in zwei Schichten ein:
Schicht 1 umfaßt die Gärten mit 42 und weniger Bäumen;
Schicht 2 setzt sich aus Gärten zusammen mit 200 und mehr Bäumen.

Die Schichtgrößen der Grundgesamtheit sind:

$N_1 = 200$ und $N_2 = 64$.

Für die einzelnen Schichten errechnen wir als Mittelwerte:

$$\bar{y}_1 = 31{,}54 \qquad \text{und} \qquad \bar{y}_2 = 239{,}81 \quad (dz)$$

Der Schätzwert der A-posteriori-Schichtung wird also:

$$\hat{Y}_p = N\bar{y}_p = \sum_h N_h \bar{y}_h = 21\,656 \text{ dz}$$

Die Schichtvarianzen sind:

$$s_1^2 = 45{,}91 \qquad \text{und} \qquad s_2^2 = 1\,924{,}92$$

Der Schätzwert der Varianz der A-posteriori-Schichtung wird dann:

$$v(\hat{Y}_p) = \frac{N-n}{n}\left[\sum_h N_h s_h^2 + \frac{1}{n}\sum_h (N-N_h)s_h^2\right] = 223\,460$$

und die Standardabweichung ergibt:

$$s_{\hat{Y}_p} = \sqrt{v(\hat{Y}_p)} = 473 \quad (dz)$$

Der wahre Wert liegt mit einer Wahrscheinlichkeit $\Phi(2)=95{,}45\%$ zwischen den Grenzen

$$20\,210 \leq Y \leq 22\,602$$

3.5.2.3 Verhältnisschätzung

Bei der Verhältnisschätzung berechnen wir zunächst das Verhältnis

$$f = \frac{\bar{y}}{\bar{x}} = 0{,}9879$$

Den Schätzwert berechnen wir dann mit

$$\hat{Y}_v = f \cdot X = 22\,209 \text{ dz}$$

Für den Schätzwert der Varianz erhalten wir:

$$v(\hat{Y}_v) = \frac{N(N-n)}{n}(s_y^2 + f^2 s_x^2 - 2fs_{xy}) = 269\,552$$

Daraus erhalten wir die Standardabweichung:

$$s_{\hat{Y}_v} = 519 \quad (dz)$$

Der wahre Wert liegt mit einer Wahrscheinlichkeit $\Phi(2)$ zwischen den Grenzen

$$21\ 171 \leqq Y \leqq 23\ 247$$

3.5.2.4 Differenzenschätzung

Bei der Differenzenschätzung nehmen wir der Einfachheit halber, und da wir keine andere Information darüber haben k=1. Damit erhalten wir als Schätzwert:

$$\hat{Y}_d = N\left[\bar{y} + (\bar{X}-\bar{x})\right] = 22\ 200 \text{ dz}$$

Der Schätzwert der Varianz wird

$$v(\hat{Y}_d) = \frac{N(N-n)}{n}\left(s_y^2 + s_x^2 - 2s_{xy}\right) = 267\ 725$$

mit der Standardabweichung

$$s_{\hat{Y}_d} = 517 \quad \text{(dz)}$$

Für t=2 ergeben sich wieder die Grenzen mit

$$21\ 146 \leqq Y \leqq 23\ 234$$

Der Unterschied in der Wirksamkeit zwischen Verhältnisschätzung und Differenzenschätzung ist natürlich keineswegs signifikant. Die theoretischen Aspekte hierzu haben wir an anderer Stelle (Abschnitt 3.5.1.5) behandelt.

3.5.2.5 Regressionsschätzung

Für die Regressionsschätzung berechnen wir zunächst den Regressionskoeffizienten

$$b = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} = 1{,}023$$

Als Schätzwert erhalten wir dann

$$\hat{Y}_r = N\left[\bar{y} + b(\bar{X}-\bar{x})\right] = 22\ 176 \text{ dz}$$

Da wir den Korrelationskoeffizienten schon am Anfang ausgewiesen haben, erhalten wir als Schätzwert der Varianz:

$$v(\hat{\bar{Y}}_r) = \frac{N(N-n)}{n} \frac{n-1}{n-2} s_y^2(1-r^2) = 221\ 902$$

Daraus ergibt sich als Standardabweichung:

$$s_{\hat{\bar{Y}}_r} = 471 \quad (dz)$$

Der wahre Wert liegt mit einer Wahrscheinlichkeit $\Phi(2)$ zwischen den Grenzen

$$21\ 234 \leqq \bar{Y} \leqq 23\ 118$$

3.5.2.6 Vergleichende Zusammenstellung

Zum Schluß wollen wir die Schätzwerte für Mittelwert, Varianz und Variationskoeffizient zur besseren Anschauung in einer Tabelle zusammenstellen. Der Variationskoeffizient ist dabei in üblicher Weise definiert als

$$c_{\hat{\bar{Y}}..} = \frac{s_{\hat{\bar{Y}}..}}{\hat{\bar{Y}}} 100$$

Die Hochrechnungsmethoden sind den Varianzen entsprechend in absteigender Reihenfolge aufgeführt.

Tabelle 2 Vergleichende Zusammenstellung

Hochrechnungsmethode	Mittelwert	Varianz	Variationskoeffizient
Einfache Hochrechnung	23 171	3 972 161	8,6%
Verhältnisschätzung	22 209	269 552	2,3%
Differenzenschätzung	22 200	267 725	2,3%
A-posteriori-Schichtung	21 656	223 460	2,2%
Regressionsschätzung	22 176	221 902	2,1%

An anderer Stelle haben wir ein ähnliches Beispiel durchge-

führt[1]. Zum Vergleich wollen wir dieses heranziehen, weil es zeigt, wie ausschlaggebend für die Wirksamkeit der A-posteriori-Schichtung die Verteilung bzw. der Schichtungseffekt ist. In diesem Beispiel umfaßte die Schichtung in Schicht 1 die Gärten mit 28-36 Bäumen und in Schicht 2 jene mit 45-68 Bäumen. Auch hier wurde wie im jetzigen Beispiel ein Ertrag pro Baum von 70 bis 130kg angenommen. Der Aufbau war also ganz ähnlich, nur daß im jetzigen Beispiel die Schichtung viel stärker und somit der Schichtungseffekt wesentlich größer war; denn Schicht 1 umfaßte die Gärten mit 24-42 Bäumen und Schicht 2 jene mit 200-290 Bäumen. Die Ergebnisse wollen wir wieder in absteigender Reihenfolge den Varianzen entsprechend ordnen.

Tabelle 2a Vergleichende Zusammenstellung eines früheren Beispiels

Hochrechnungsmethode	Mittelwert	Varianz	Variationskoeffizient
Einfache Hochrechnung	10 039	37 148	1,92%
A-posteriori-Schichtung	9 822	8 559	0,95%
Verhältnisschätzung	9 765	1 397	0,38%
Differenzenschätzung	9 760	1 367	0,38%
Regressionsschätzung	9 757	1 239	0,36%

Wir sehen aus diesen beiden Zusammenstellungen, daß die Reihenfolge in der Wirksamkeit der Schätzverfahren mit Ausnahme der A-posteriori-Schichtung gleich geblieben ist. Hier wird deutlich demonstriert, wie ausschlaggebend allein der Schichtungseffekt für die Wirksamkeit dieses Verfahrens ist.

[1]) Hauser, Siegfried (1965)(unveröff.): S.159ff..

3.6 Schichtenbildung

Wie wir schon an früherer Stelle (s. Abschnitt 3.1 Prinzip dieser Methode) bemerkt haben, gilt als Vorzug der geschichteten Verfahren ihr praktisch von vornherein garantierter Gewinn an Genauigkeit. Jedoch hängt die Größe dieser Effizienz - hier speziell bei der A-posteriori-Schichtung, wie wir auch im vorhergehenden Abschnitt (3.5 Wirksamkeit) deutlich gesehen haben - vom Schichtungseffekt ab. Die Größe des Schichtungseffekts hängt von der mehr oder weniger starken Schichtung der Merkmale in der Grundgesamtheit ab und davon, wie gut diese Schichtgrenzen bei der Schichtung getroffen werden.

In einem ersten Abschnitt wollen wir die Theorie - man könnte sie fast schon konventionell nennen - behandeln, die mit den Namen Neyman und Dalenius aufs engste verbunden ist. Man kann diese Theorie auch als einen Bestandteil der Vorbereitung des Stichprobenplans verstehen, im Sinne einer Aufteilung der Grundgesamtheit[1]. Aus diesem Grunde - da wir es hier mit einem Hochrechnungsverfahren zu tun haben - und weil diese Theorie in größerem Maße für das geschichtete Auswahlverfahren (A-priori-Schichtung) Bedeutung hat, wollen wir dieses Problem hier nicht ausführlich behandeln und auf die reichliche Literatur verweisen. Wir wollen sie die allgemeine Theorie der Schichtung nennen.

In einem zweiten Abschnitt wollen wir das für ein Hochrechnungsverfahren wichtigere Thema behandeln, nämlich wie die als zusätzliche Information benötigten Gewichte für die A-posteriori-Schichtung geschätzt werden können. Diese Problematik wollen wir als Schätztheorie der Schichtung bezeichnen.

3.6.1 Allgemeine Theorie der Schichtung

Obwohl die Probleme der Schichtung in einem erheblichen Umfang in den Lehrbüchern und Zeitschriften behandelt werden, kann man

---

[1]) Vgl. Mahalanobis, P.C. (1952): S.4.

nicht von einer abgeschlossenen Theorie sprechen. Dabei kann das Jahr 1950 - ausgelöst durch einen Artikel von Dalenius[1] - als der Beginn einer verstärkten Diskussion dieser Thematik betrachtet werden.

Die Aufteilung in Schichten hängt von der Häufigkeitsverteilung der Variablen ab. Dabei kommt der Wirksamkeit der geschichteten Verfahren zugute, daß wir es in der Wirtschafts- und Sozialstatistik oft mit Verteilungen zu tun haben, bei denen sich gut Schichten herausbilden lassen. Solche - meist schiefen Verteilungen - finden wir z.B. im Schul- und Krankenhauswesen, bei Unternehmensgrößen, Umsatzhöhen, bei der Einkommensteuer bzw. dem Einkommen.[2] Die Schichtung kann vorgenommen werden bezüglich des untersuchten Merkmals oder bezüglich eines diesem eng korrelierten Merkmals. Z.B. wird der Getreideertrag untersucht, aber geschichtet wird nach der Anbaufläche.

Zur Schichtenbildung finden wir in der Literatur sehr nützliche, allerdings meist nur verbale, Prinzipien.[3] Es wird dabei gesagt, die Schichten sollen so homogen wie möglich sein, und die untersuchte Variable soll mit der Schichtungsvariablen eng korreliert sein. Auf der anderen Seite sollen die Mittelwerte sehr heterogen sein, was bedeutet, daß die Variable in der Größe erheblich variieren muß; zu kleine Schichten tragen also wenig zur Wirksamkeit bei. Gering besetzte Schichten sollen voll erhoben werden. Bei Einführung neuer Schichten soll der zusätzliche Genauigkeitsgewinn geprüft werden.

---

1) Dalenius, Tore (1950): S.203-213.

2) Vgl. Cochran, William G. (1964): S.100; Hess, Irene/Sethi, V.K./Balakrishnan, T.R. (1966): S.74.

3) Vgl. z.B. Cochran, William G. (1964): S.100; Deming, W. Edwards (1960): S.487ff.; Désabie, J. (1966): S.137ff.; Kellerer, Hans (1963): S.100; Kish, Leslie (1965): S.100; Stephan, Frederick F./McCarthy, Philip J. (1958): S.415. Interessant ist hier, daß auch die Biostatistiker die Probleme der Schichtung und Klassifizierung noch nicht gelöst haben, obzwar man meinen könnte, dort seien die Probleme einfacher, da es sich meist um natürliche Klassifizierungsmerkmale handelt. Doch auch dort wird geklagt, daß man in den Methoden zur Klassenbildung "seit hundert oder gar zweihundert Jahren" wenig vorangekommen ist (s. Sokal, Robert R./Sneath, Peter H.A. (1963): S.VII/VIII), obwohl so bedeutende Statistiker wie R.A. Fisher, Pearson, Mahalanobis und Rao an diesen Problemen gearbeitet haben (ebenda S.38).

Die Probleme der Schichtenbildung sind gelöst bei einer natürlichen Schichtung (z.B. Geschlecht), oft auch bei einer geografischen Schichtung. Es muß aber beachtet werden, daß eine geografische Schichtung nur dann einen Sinn hat, wenn sie mit dem untersuchten Merkmal korreliert ist, und somit auch Aussagekraft für es besitzt.

Ein anderes Schichtproblem, die Mehrfachschichtung (multiple oder deep stratification)[1], wollen wir hier noch streifen. In der Literatur wird sie nur kurz und meist verbal behandelt. Dieses Problem ist noch nicht gelöst und enthält eine Reihe neuer Ansatzpunkte für die Auswahl- und Hochrechnungsmethoden. Mit Hilfe der Mehrfachschichtung kann eine Anzahl von Schichten ausgewählt werden, die eine weit größere Anzahl von Schichten repräsentiert.[2]

Neben diesen verbalen Prinzipien finden wir in der Literatur Formeln, die es ermöglichen, die Schichten rechnerisch festzulegen. Zunächst handelt es sich um die Aufteilung der Anzahl der Stichprobeneinheiten auf die einzelnen Schichten (Allokationsformen). Das Problem der optimalen Allokation, bei dem die Kosten mitberücksichtigt werden, wollen wir hier außer Acht lassen, weil sich dadurch keine grundsätzlich andere Problemstellung ergibt. Die bekanntesten Allokationsformen sind:[3]

a) gleiche Aufteilung auf die Schichten

$$n_h = \frac{n}{L}$$

b) proportionale Aufteilung

$$n_h = n \frac{N_h}{N}$$

c) proportionale Aufteilung zu den Schichtaggregaten der zu schichtenden Variablen. Hier würde z.B. nicht proportional zu den Betten von Krankenhäusern ausgewählt, wie dies bei b) der Fall wäre, sondern zur durchschnittlichen Anzahl der Betten aller Kranken-

---

1) Vgl. Sukhatme, P.V. (1966): S.373.

2) Vgl. Tepping, Benjamin J./Hurwitz, William N./Deming, W. Edwards (1943): S.93 und Bryant, Edwards C./Hartley, H.O./Jessen, R.J. (1960): S.105-124.

3) Vgl. Hess, Irene/Sethi, V.K./Balakrishnan, T.R. (1966): S.75.

häuser:

$$n_h = n \frac{N_h \mu_h}{\Sigma N_h \mu_h}$$

d) die Aufteilung nach Neyman, mit

$$n_h = n \frac{N_h S_h}{\Sigma N_h S_h}$$

Die proportionale Aufteilung hat den Vorteil, daß man die Parameter S_h zur Schichtaufteilung nicht kennen muß[1]. Da die Standardabweichungen der Schichten in der Grundgesamtheit oft nicht bekannt sind, versucht man diesen Nachteil der Neyman-Allokation dadurch auszugleichen, daß man nach Verfahren sucht, mit Hilfe derer man diese Parameter schätzen kann, ohne daß größere Genauigkeitsverluste eintreten[2].

Suchen wir nun auf Grund der angegebenen Aufteilungsformen die besten (optimalen) Schichtungspunkte y_h, so finden wir, daß in den 4 Fällen folgende Bedingungen erfüllt sein müssen, bei

a)[3]
$$y_h = \frac{\bar{Y}_h + \bar{Y}_{h+1}}{2}$$

$$(h=1,2,\ldots,L-1)$$

b)[4]
$$\frac{N_h}{N}\left[S_h^2 + (y_h - \bar{Y}_h)^2\right] = \frac{N_{h+1}}{N}\left[S_{h+1}^2 + (y_h - \bar{Y}_{h+1})^2\right]$$

c)[5]
$$\frac{\bar{Y}_h(y_h - \bar{Y}_h)^2 - S_h^2(y_h - \bar{Y}_h)}{\bar{Y}_h^2} = \frac{\bar{Y}_{h+1}(y_h - \bar{Y}_{h+1})^2 - S_{h+1}^2(y_h - \bar{Y}_{h+1})}{\bar{Y}_{h+1}^2}$$

d)[6]
$$\frac{(y_h - \bar{Y}_h)^2 + S_h^2}{S_h} = \frac{(y_h - \bar{Y}_{h+1})^2 + S_{h+1}^2}{S_{h+1}}$$

---

1) Beispiele für die proportionale und Neyman-Allokation sind zu finden bei: Strecker, Heinrich (1957): S.80ff..

2) S. Rao, T.J. (1968): S.159ff..

3) Vgl. Hess, Irene/Sethi, V.K./Balakrishnan, T.R. (1966): S.88.

4) Cochran, William G. (1961): S.354.

5) Vgl. Hess, Irene/Sethi, V.K./Balakrishnan, T.R. (1966): S.88.

6) Vgl. Kish, Leslie (1965): S.104.

Für die Fälle b, c und d sind die optimalen Lösungen nicht möglich, da alle Parameter in den Gleichungen von y_h abhängen.[1)]

Daher wurde nach praktischen Anwendungsregeln gesucht, die Werte liefern, die den angegebenen Optima nahe kommen. Solche Näherungsformeln werden angegeben von

Ekman[2)] mit $\frac{N_h}{N}(y_h - y_{h-1}) = \text{const.}$, von

Mahalanobis[3)] mit $\frac{N_h}{N}\bar{Y}_h = \text{const.}$ und von

Dalenius/Gurney[4)] mit $\frac{N_h}{N}S_h = \text{const.}$.

Dalenius und Hodges[5)] konstruieren die Schichtintervalle, indem die Quadratwurzeln der Häufigkeiten der Schichtungsvariablen aufsummiert werden, und die Summe in angenähert gleiche Teile geteilt wird. Weitere Näherungsformeln, um die Schichtungspunkte zu finden, werden gegeben von Aoyama[6)], Durbin[7)] und Sethi[8)]. Aus der Fülle der angegebenen Formeln, die gleichwertig nebeneinanderstehen, ist schon zu ersehen, daß eine umfassende Theorie hier erst im Entstehen ist. Ein Vergleich der verschiedenen Allokations- und Konstruktionsformen für die Schichten wird dann zeigen, welche Formeln für die praktische Stichprobenerhebung von Bedeutung sein werden[9)].

Allgemein gilt für die Schichtanzahl L_1 und L_2:

Ist $L_1 > L_2$, so bedeutet das $\sigma^2(x;L_1) < \sigma^2(x;L_2)$.

---

1) Vgl. Kish, Leslie (1965): S.104.
2) Ekman, G. (1959): S.219-229.
3) Mahalanobis, P.C. (1952): S.1-7.
4) Dalenius, Tore/Gurney, M. (1951): S.133-148.
5) Dalenius, Tore/Hodges, Josef L. (1959): S.88-101.
6) Vgl. Dalenius, Tore/Hodges, Josef L. (1957): S.199.
7) Durbin, J. (1959a): S.246-248.
8) Sethi, V.K. (1963): S.20-33.
9) Hier wurde ein Anfang gemacht von Hess, Irene/Sethi, V.K./Balakrishnan, T.R. (1966): S.82/83.

Diese allgemeine und triviale Aussage führt den Praktiker nicht weiter, weil damit noch nichts darüber gesagt ist, wie sich der Genauigkeitsgewinn bei Vergrößerung der Zahl der Schichten zur Zunahme des Arbeitsaufwandes bzw. der Kosten verhält; denn sind mit Hilfe der oben angegebenen Näherungsformeln die Schichtungspunkte und die Anzahl der Schichten festgelegt, so ist nicht sicher, ob im Vergleich mit den entstehenden Kosten eine optimale Schichtung vorliegt. Kish[1] meint hier - wohl aus praktischen Stichprobenerfahrungen heraus und weniger theoretisch fundiert - die optimale Schichtanzahl sei 3-6 Schichten. Es ist natürlich möglich, mit Hilfe der "cut-and-try - Methode" in der einen Schicht den Stichprobenumfang zu vergrößern und in der anderen zu verkleinern.[2] Man kann hier - etwas umständlich - zu einer unter Kostengesichtspunkten optimalen Schichtaufteilung kommen. Eine Aussage über eine effiziente Schichtung ist nur durch den Vergleich von Genauigkeitszuwachs und zusätzlichen Kosten möglich. Die Optimierungsprobleme sehen zunächst ähnlich aus wie jene, die aus der Wirtschaftstheorie bekannt sind. Wir können hier analog von folgenden Kurven ausgehen (vgl. dazu S.28):

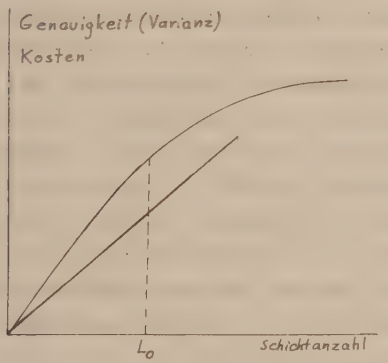


Abb.12a Vergleich von Genauigkeit und (linearen) Kosten bei unterschiedlicher Schichtanzahl

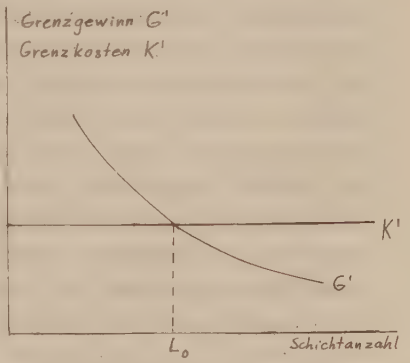


Abb.12b Vergleich von zusätzlichem Genauigkeitsgewinn und zusätzlichen Kosten (Grenzbetrachtung)

---

[1]) Kish, Leslie (1965): S.102.

[2]) Deming, William Edwards (1966): S.213.

Die beiden Schaubilder zeigen, wie erwartet, daß die optimale Schichtanzahl L_o dort ist, wo der zusätzliche Genauigkeitsgewinn gleich den zusätzlichen Kosten ist. Das bedeutet aber, daß nur eine empirische Betrachtung der Verteilung der Merkmale und der Kostenkurve hier eine Lösung bringt[1]. Die Berechnung erfordert darüber hinaus Methoden zur Lösung dynamischer Optimierungsprobleme wegen des Index'. Bis jetzt wird in der Literatur meist auf die alte Faustregel hingewiesen, daß wenig Schichten das meiste an Genauigkeitsgewinn bringen; weitere Schichten bringen wenig zusätzlichen Gewinn[2].

3.6.2 Schätztheorie der Schichtung

Für die A-posteriori-Schichtung bedeutsamer als die im letzten Abschnitt behandelte ist eine Theorie, auf Grund derer die Schichtanteile geschätzt werden. Die zusätzliche Information ist in diesem Fall bei Vorliegen der Daten nicht gegeben, und sie soll rasch und doch in ausreichendem Maße gesichert gefunden werden. Bekanntlich ist bei der Beurteilung von Stichprobenergebnissen von zwei Forderungen auszugehen: von der Genauigkeit und der Aktualität der Ergebnisse. Neben der Genauigkeit wird der Zeitfaktor nicht immer genügend berücksichtigt. Auf diesem Gebiet liegt dann auch die Bedeutung der Markt- und Meinungsforschung, die meist ohne Angabe der Genauigkeit, jedoch mit großer Aktualität ihre Ergebnisse liefert[3]. Hier sollte weitgehend die theoretische Statistik Methoden liefern, die bei angebbarer Genauigkeit genauso aktuelle Ergebnisse zu liefern vermögen. Aus dieser Sicht sollten solche Hochrechnungsmethoden verfeinert bzw. neu entwickelt werden, die bei geringem Zeitmehraufwand die Genauigkeit erheblich steigern und dem Praktiker dennoch handliche Formeln liefern. Gelingt es, für die Schichtanteile gute Schätzwerte zu finden, so könnte die A-po-

---

1) Vgl. hierzu Désabie, J. (1966): S.140ff., dazu Dalenius, Tore (1957): S.117, der eine Untersuchung bei drei verschiedenen Häufigkeitsverteilungen vorgenommen hat.

2) S. z.B. Kish, Leslie (1965): S.102.

3) Mit Hilfe eines Telefon-Panels können schon in wenigen Stunden Ergebnisse einer Umfrage geliefert werden.

steriori-Schichtung die geforderten Bedingungen einer großen Genauigkeit und Aktualität erfüllen.

3.6.2.1 Einfache Schätzung aus der Stichprobe

Bisher gingen wir von bekannten Schichtanteilen N_h aus. Bei der A-posteriori-Schichtung brauchen wir zur Schätzung dieser Größen - im Gegensatz zum geschichteten Auswahlverfahren (A-priori-Schichtung) - auf Kostengesichtspunkte keine Rücksicht zu nehmen[1]; denn die Auswahl mit den damit verbundenen Kosten ist durchgeführt. Bei der Schätzung der Schichtanteile N_h/N sollen diese Schätzwerte der zugrunde liegenden (unbekannten) Häufigkeitsverteilung gut angepaßt werden, damit ein etwa auftretender Fehler gering gehalten wird.

Will man die Schichtanteilswerte nicht aus einer anderen - jedoch mit der untersuchten gut korrelierten - Grundgesamtheit übernehmen oder sie durch Plausibilitätsüberlegungen irgendwie finden[2], bleibt als wichtigste Möglichkeit, sie aus der vorliegenden Stichprobe zu schätzen.

Als Schätzwert für die Schichtanteile nehmen wir:

$$\left(\frac{\widehat{N_h}}{N}\right) = \frac{r_h}{n}$$

wobei

$$E\,\frac{n_h}{n} = \frac{N_h}{N}$$

ein unverzerrter Schätzwert ist.

Die Schätzfunktion, die uns die Werte bei der A-posteriori-Schichtung liefert, ist dann

$$\hat{\bar{y}}_p = \sum_{h=1}^{L} \frac{n_h}{n}\,\bar{y}_h \qquad (30)$$

Bei gegebenen n_h ist der Erwartungswert (bedingter Erwartungswert) dieser Schätzfunktion:

---

1) Vgl. Deming, Edwards W. (1960): S.321.

2) Diese Möglichkeiten können nicht allgemeintheoretisch, sondern nur aus dem Einzelfall heraus betrachtet und begründet werden.

$$E(\hat{\bar{y}}_p | n_h) = \Sigma \frac{n_h}{n} E\bar{y}_h = \Sigma \frac{n_h}{n} \bar{Y}_h$$

Der unbedingte Erwartungswert wird dann:

$$E(\hat{\bar{y}}_p) = E\Sigma \frac{n_h}{n} \bar{Y}_h = \Sigma \bar{Y}_h E \frac{n_h}{n} = \Sigma \frac{N_h}{N} \bar{Y}_h = \bar{Y}$$

Die Schätzfunktion liefert uns also erwartungstreue Schätzwerte.

Die Varianz dieses Schätzwertes ist:

$$V(\hat{\bar{y}}) = \frac{N-n}{N \cdot n} S^2$$

d.h. die Varianz der freien Hochrechnung. Der Beweis läßt sich am einfachsten führen, wenn zwei Schichten angenommen werden, deren Anteile in der Stichprobe n_1/n und n_2/n betragen. Dann gilt:

$$V(\hat{\bar{y}}_p) = E(\hat{\bar{y}}_p - \bar{Y})^2 = E(\frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n} - \bar{Y})^2$$

$$= E(\bar{\bar{y}} - \bar{Y})^2 = \frac{N-n}{N \cdot n} S^2$$

Bei dieser Methode haben wir keinen Genauigkeitsgewinn gegenüber der freien Hochrechnung zu verzeichnen. Ihre Anwendung hat nur Sinn, wenn wir die Schichten nachträglich zur Gewinnung von Aussagen bilden müssen, da die Zugehörigkeit zu den Schichten nur nachträglich festgestellt werden kann (z.B. Religionszugehörigkeit, Parteizugehörigkeit, Konsumgewohnheiten) und die Schichtanteilssätze in der Grundgesamtheit nicht bekannt sind.

Wenn bei fortlaufenden Stichproben über denselben Untersuchungsgegenstand die Schätzwerte n_h/n eine deutliche Konstanz in der Zeit aufzeigen, nehmen diese Schätzwerte immer mehr den Charakter einer "wahren" Information an[1]. Die Schätzwerte der Anteilssätze können dann in die Varianzformel der A-posteriori-Schichtung als "wahre" Anteilssätze eingesetzt werden, so daß wir die

---

[1]) Hier tritt das auf, was Mahalanobis, P.C. (1952): S.5 mit "sequential (historical) designs of sample surveys" bezeichnet, nämlich der Vorgang, daß die in der Zeit sich anhäufenden Informationen dazu benutzt werden, die folgenden Stichprobenpläne zu verbessern.

größere Genauigkeit der A-posteriori-Schichtung erhalten, obwohl die Anteilssätze aus der Stichprobe geschätzt sind.

Eine andere - hier nur kurz gestreifte - Methode zur Schätzung der Schichtanteile zeigt Sukhatme auf[1]. Dabei wird aus einer vorläufigen Stichprobe auf Grund einer Unterstichprobe der Charakter der Verteilung untersucht. Die Grundgesamtheit wird dann proportional oder nach der von Neyman vorgeschlagenen Methode unter Berücksichtigung der Kosten aufgeteilt. Diese Methode, die für das geschichtete Auswahlverfahren (A-priori-Schichtung) vorgeschlagen wird, ist im Prinzip auch für die A-posteriori-Schichtung anwendbar, indem hier nachträglich der Charakter der Verteilung untersucht wird. Sie ist jedoch umständlicher und wesentlich zeitraubender als die oben behandelte. Ferner fallen die Kostengesichtspunkte bei der A-posteriori-Schichtung nicht ins Gewicht.

Eine ähnliche Problemstellung wie die behandelte ergibt sich für die Non-response-Rate bei Stichprobenerhebungen. Dort wird versucht, über die Gruppe der Nichtbeantworter mit Hilfe einer Unterstichprobe Aussagen zu erhalten. Zur Berechnung der Ergebnisse der Gesamtstichprobe müßte die Non-response-Rate ($N_{\overline{A}}$) bzw. die Rate der Beantworter (N_A) in der Grundgesamtheit bekannt sein. Diese unbekannten Anteilssätze werden mit Hilfe der Stichprobengrößen $n_{\overline{A}}$ und n_A geschätzt. Da wir aber für die Non-response-Rate nur eine Unterstichprobe zur Verfügung haben, wird die Varianz des Schätzwertes der Non-response-Rate größer als der Ausdruck für die einfache Hochrechnung, d.h., zur Varianz der einfachen Hochrechnung muß noch ein Faktor hinzugefügt werden[2].

3.6.2.2 Schätzung der Anteilssätze unter Annahme einer A-priori-Verteilung

Wir schätzen die $P_h = N_h/N$, indem wir über die Verteilungsform einige Annahmen machen. Man kennt also die Parameter nicht,

---

[1] Sukhatme, Pandurang V. (1963): S.112ff..

[2] Vgl. Sukhatme, Pandurang V. (1963): S.481.

geht aber davon aus, daß deren Verteilungsform bekannt sei[1]. Als Verteilungsform dieser Variablen, die Werte zwischen 0 und 1 annimmt, nehmen wir eine Betaverteilung an. Die Betaverteilung läuft von 0 bis 1, und ebenso können die Anteilssätze der Schichten nur Werte von 0 bis 1 annehmen. Hendricks meint, derselbe Gedankengang, der den Statistiker bei der Anwendung Bayes'scher Methoden zur Binomialverteilung führt, veranlaßt ihn "seit undenklichen Zeiten", zur Beschreibung einer solchen Verteilung eine Beta-Funktion anzunehmen[2].

Die Betaverteilung wird in jüngster Zeit auf dem Gebiet des Operational Research bei Netzwerktechniken angewandt, z.B. beim Planungssystem PERT (Programm Evaluation and Review Technique). Das Netzwerkmodell PERT wurde 1958 im Zusammenhang mit dem Polaris-Raketen-Programm entwickelt. Dieses Planungssystem wird heute bei Planungs-, Entwicklungs-, Konstruktions- und Produktionsmodellen im Militärbereich, in der Industrie, aber auch auf dem kriminalistischen Sektor (Nachbildung des englischen Postraubs 1963 durch die Polizei zur Aufdeckung des Verbrechens) und zur Planung des Schulunterrichts eingesetzt[3]. PERT ist im Gegensatz zu anderen Netzwerkmodellen, z.B. CPM (Critical Path Method) kein streng determiniertes Modell, sondern ein stochastisches Modell[4].

Für die Wahl der Betaverteilung zur Schätzung der Parameter P_h, wenn sonst keine Informationen vorliegen, sprechen einige Argumente[5]. Die Betaverteilung ist

- unimodal
- stetig
- sie besitzt zwei nicht-negative Abszissenendpunkte
- die Betaverteilung ist in gewisser Weise die Verallgemeinerung der Gleichverteilung[6] (das sieht man, wenn

---

1) Vgl. Jorgenson, D.W./McCall, J.J./Radner, R. (1967): S.101.

2) Hendricks, Walter A. (1964): S.227ff.. Daß die Beta-Funktion zur Anpassung an bestimmte Verteilungen gute Werte erbringt, zeigt ein Beispiel bei David, F.N./Barton, D.E. (1962): S.114ff..

3) S. Schreiter, Dieter/Stempell, Dieter (1968): S.8-11.

4) Ebenda S.7.

5) Vgl. McCrimmon, Kenneth R./Ryavec, Charles A. (1966): S.523/524.

6) S. Ferguson, Thomas S. (1967): S.101/102.

man im folgenden bei (31) a=b=1 bzw. bei (34) a=1 setzt)

- die Betaverteilung geht bei größer werdendem Stichprobenumfang in die Binomialverteilung und weiter in die Normalverteilung über[1].

Die angeführten Gründe für die Anwendung der Betaverteilung als A-priori-Verteilung zum Schätzen von Parametern lassen diese beinahe als Idealverteilung erscheinen, da sie eine Reihe von Verteilungstypen als Grenzverteilungen besitzt. Es ist natürlich möglich, daß man mit der Wahl der Beta-Funktion eine Verzerrung bei der Schätzung erhält, wenn die tatsächliche Verteilung durch die Betaverteilung nicht approximiert wird[2]. Deshalb sollten hier zunächst genaue empirische Untersuchungen stattfinden, damit nicht ein ähnlicher Streit entsteht, wie in den Fällen, in denen oft die Normalverteilung zugrundegelegt wird, wo dann diskutiert wird, ob diese ein "Naturgesetz oder Ideologie" ist[3].

Bei der Behandlung der Betaverteilung wollen wir uns auf die mathematischen Herleitungen beschränken, die unser Spezialfall hier benötigt[4]. Die Beta-Funktion (Dichtefunktion) ist allgemein definiert für eine Variable z, die die Werte von 0 bis 1 annimmt[5]:

$$\beta(z) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} z^{a-1}(1-z)^{b-1}$$

$$= \frac{(a+b-1)!}{(a-1)!(b-1)!} z^{a-1}(1-z)^{b-1} \qquad (31)$$

---

1) Die mathematischen Herleitungen sind zu finden bei: Peters, William S./Summers, George W. (1968): S.203/204 und Moran, P.A.P. (1968): S.328-332.

2) Über Fehler und die Größe der Fehler, die durch die Zugrundelegung der Beta-Funktion entstehen können, s.: McCrimmon, Kenneth R./Ryavec, Charles A. (1966): S.525-528.

3) Siehe die Ausführungen bei Menges, Günter (1968): S.248/249.

4) Dazu sei verwiesen auf Wadsworth, George P./Bryan, Joseph G. (1960): S.4-6 und S.101-104 und Mauldon, J.G. (1959): S.509-520.

5) Vgl. Karlin, Samuel (1966): S.8;
Smirnow, W.I. (1959): S.223ff..

wobei beide Parameter a und b größer als 0 sein müssen. Der Mittelwert einer solchen Verteilung ist:

$$\bar{z} = \frac{a}{a+b}$$

Das 2.Moment um 0:

$$Ez^2 = \frac{a+1}{a+b+1} \frac{a}{a+b}$$

Somit wird die Varianz um den Mittelwert $\bar{z}$:

$$V(\bar{z}) = Ez^2 - \bar{z}^2 = \frac{a+1}{a+b+1} \frac{a}{a+b} - \left(\frac{a}{a+b}\right)^2$$

$$= \frac{a}{a+b}\left(\frac{a+1}{a+b+1} - \frac{a}{a+b}\right)$$

$$= \frac{a \cdot b}{(a+b)^2(a+b+1)} \qquad (32)$$

Aus der Stichprobe haben wir die Schätzwerte p_h vorliegen, darüber hinaus liegen aus der Stichprobe keine Informationen über die Verteilung vor. Wir suchen nun den Erwartungswert von P_h bei einem gegebenen Satz von n_h und nehmen diesen als Schätzwert $\hat{p}_h$.

Die P_h können nur Werte zwischen 0 und 1 annehmen. Der Mittelwert aller P_h-Werte über L Schichten wird:

$$\mu = \frac{\sum_h^L P_h}{L} = \frac{1}{L}$$

Der Mittelwert bei der Betaverteilung war

$$\bar{z} = \frac{a}{a+b}$$

somit können wir b eliminieren, denn

$$\frac{1}{L} = \frac{a}{a+b}$$

gibt

$$b = a(L-1) \qquad (33)$$

In der Dichtefunktion können wir b durch a und L ersetzen. Um die Schätzwerte zu erhalten, bilden wir den bedingten Erwartungswert von P_h bei einer gegebenen Stichprobe beobachteter Werte n_h.

Wir nehmen als Dichtefunktion für die P_h bei gegebenen n_h [1]:

$$f(P_h) = \frac{\Gamma(n+La)}{\Gamma(n_1+a)\Gamma(n_2+a)\ldots\Gamma(n_L+a)} P_1^{n_1+a-1} P_2^{n_2+a-1} \ldots P_L^{n_L+a-1} \qquad (34)$$

Somit wird der bedingte Erwartungswert der P_h:

$$E(P_h|n_h) = \underbrace{\int_0^1 \int_0^1 \ldots \int_0^1}_{L-1} P_h f(P_h) dP_1 dP_2 \ldots dP_{L-1}$$

$$= \frac{\Gamma(n+La)}{\Gamma(n_1+a)\Gamma(n_2+a)\ldots\Gamma(n_L+a)} \underbrace{\int_0^1 \int_0^1 \ldots \int_0^1}_{L-1} P_h \cdot P_1^{n_1+a-1} \ldots P_h^{n_h+a-1} \ldots$$
$$\ldots P_L^{n_L+a-1} dP_1 dP_2 \ldots dP_{L-1}$$

$$= \frac{\Gamma(n+La)}{\Gamma(n_1+a)\Gamma(n_2+a)\ldots\Gamma(n_L+a)} \underbrace{\int_0^1 \int_0^1 \ldots \int_0^1}_{L-1} P_1^{n_1+a-1} \ldots P_h^{n_h+a} \ldots$$
$$\ldots P_L^{n_L+a-1} dP_1 dP_2 \ldots dP_{L-1}$$

$$= \frac{\Gamma(n+La)\left[\Gamma(n_1+a)\ldots\Gamma(n_h+a+1)\ldots\Gamma(n_L+a)\right]}{\Gamma(n_1+a)\Gamma(n_2+a)\ldots\Gamma(n_L+a)\left[\Gamma(n+La+1)\right]}$$

$$= \frac{\Gamma(n+La)\Gamma(n_h+a+1)}{\Gamma(n_h+a)\Gamma(n+La+1)}$$

$$= \frac{n_h+a}{n+La}$$

Diesen Erwartungswert nehmen wir als Schätzwert und bekommen für $\hat{p}_h$ den einfachen Ausdruck:

---

[1]) Vgl. Hendricks, Walter A. (1964): S.226ff..

$$\hat{p}_h = \left(\widehat{\frac{n_h}{n}}\right) = \frac{n_h+a}{n+La} \qquad (35)$$

Um diesen Schätzwert berechnen zu können, benötigen wir den Parameter a. Wir suchen also eine Gleichung für a und bilden dazu die Varianz σ_p^2. Dabei gehen wir von Gleichung (32) aus, der Varianz der Betaverteilung.

Die Varianz um den Mittelwert 1/L beträgt, wenn b mit Hilfe der Beziehung (33) eliminiert wird:

$$\sigma_p^2 = \frac{a+1}{a+aL-a+1}\,\frac{1}{L} - \frac{1}{L^2} = \frac{a+1}{L(aL+1)} - \frac{1}{L^2} \qquad (36)$$

Daraus erhalten wir für

$$a = \frac{L-1-L^2\sigma_p^2}{L^3\sigma_p^2} \qquad (37)$$

Den Parameter a können wir aus der Stichprobe nur berechnen, wenn wir einen Schätzwert für σ_p^2 finden. Der Varianzausdruck (36) ist nun identisch mit folgendem Ausdruck, für den wir dann einen Schätzwert suchen:

$$\sigma_p^2 = E\,\frac{\sum_h^L (P_h - \frac{1}{L})^2}{L} \qquad (38)$$

Dabei sind die P_h die Anteilssätze in der Grundgesamtheit, für die wir die Schätzwerte finden müssen. Dazu nehmen wir einige Umformungen vor; es gilt:

$$\frac{1}{L}\sum_h^L (P_h - \frac{1}{L})^2 = \frac{1}{L}\sum_h^L (p_h - P_h + P_h - \frac{1}{L})^2$$

$$= \frac{1}{L}\,\Sigma\left[(p_h-P_h)^2 + 2(p_h-P_h)(P_h-\tfrac{1}{L}) + (P_h-\tfrac{1}{L})^2\right]$$

$$= \frac{1}{L}\,\Sigma\left[(p_h-P_h)^2 + 2(p_h-P_h)(P_h-\tfrac{1}{L})\right] + \frac{\Sigma(P_h-\frac{1}{L})^2}{L}$$

$$= \frac{1}{L}\Sigma'(p_h^2 - 2p_h P_h + P_h^2 + 2p_h P_h - 2p_h\frac{1}{L} - 2P_h^2 + 2P_h\frac{1}{L}) + \frac{\Sigma'(P_h - \frac{1}{L})^2}{L}$$

$$= \frac{1}{L}\Sigma'(p_h^2 - P_h^2) + \frac{1}{L}\Sigma'(P_h - \frac{1}{L})^2$$

weil $E\Sigma(p_h - P_h) = 0$ und damit $2\frac{1}{L}E\Sigma(p_h - P_h)$ später wegfällt. Also

$$\frac{1}{L}\Sigma(P_h - \frac{1}{L})^2 = \frac{1}{L}\Sigma(p_h - \frac{1}{L})^2 - \frac{1}{L}\Sigma(p_h^2 - P_h^2)$$

Somit erhalten wir für (38) den Ausdruck:

$$\sigma_p^2 = E\frac{\Sigma'(P_h - \frac{1}{L})^2}{L} = E\frac{\Sigma(p_h - \frac{1}{L})^2}{L} - E\frac{\Sigma'(p_h^2 - P_h^2)}{L}$$

$$= E\frac{\Sigma(p_h - \frac{1}{L})^2}{L} - \frac{1}{L}\Sigma\,\mathrm{Var}(P_h)$$

$$= E\frac{\Sigma(p_h - \frac{1}{L})^2}{L} - \frac{1}{L}\frac{\Sigma P_h(1-P_h)}{n} \qquad (39)$$

Die Varianz der Schichtanteilssätze σ_p^2 ist also gleich der Gesamtvarianz minus der Varianz innerhalb der Schichten, sofern in den Schichten eine Binomialverteilung angenommen wird. Wir haben hier so etwas ähnliches wie beim Schichtungseffekt vorliegen, wo auch von der Gesamtvarianz ein Varianzausdruck abgezogen wird.

Für die Ausdrücke auf der rechten Seite der Gleichung suchen wir nun die Schätzwerte und erhalten dann ein Schätzwert für die Varianz der Schichtanteilssätze $\hat{\sigma}_p^2$:

$$\hat{\sigma}_p^2 = \frac{\Sigma(p_h - \frac{1}{L})^2}{L} - \frac{\Sigma p_h(1-p_h)}{L(n-1)}$$

$$= \frac{1}{L}\left[\Sigma p_h^2 - 2\Sigma p_h\frac{1}{L} + \Sigma(\frac{1}{L})^2 - \frac{\Sigma p_h}{n-1} + \frac{\Sigma p_h^2}{n-1}\right]$$

$$= \frac{1}{L}\left[\frac{(n-1)\Sigma p_h^2}{n-1} - \frac{2}{L} + \frac{1}{L} - \frac{\Sigma p_h}{n-1} + \frac{\Sigma p_h^2}{n-1}\right]$$

$$= \frac{1}{L}\left[\frac{n\Sigma p_h^2}{n-1} - \frac{1}{L} - \frac{\Sigma p_h}{n-1} + \Sigma p_h - \Sigma p_h\right]$$

$$= \frac{1}{L}\left[\frac{n\Sigma p_h^2}{n-1} - \frac{1}{L} + 1 - \frac{\Sigma p_h}{n-1} - \frac{(n-1)\Sigma p_h}{n-1}\right]$$

$$= \frac{1}{L}\left[1 - \frac{1}{L} - \frac{n\Sigma p_h}{n-1} + \frac{n\Sigma p_h^2}{n-1}\right]$$

Als endgültigen Ausdruck erhalten wir:

$$\hat{\sigma}_p^2 = \frac{L-1}{L^2} - \frac{n\sum_h^L p_h(1-p_h)}{L(n-1)} \qquad (40)$$

Mit Hilfe dieses Schätzwertes[1] kann der Parameter a aus der Stichprobe geschätzt werden; dieser führt dann weiter zum gesuchten Schätzwert der Anteilssätze $\hat{p}_h$.

Hendricks[2] bringt zur Illustration ein Beispiel. Bei einer Stichprobe vom Umfang n=100 sollen die Merkmalswerte in L=3 Schichten fallen mit $p_1=n_1/n=60/100$, $p_2=n_2/n=30/100$ und $p_3=n_3/n=10/100$. Daraus ergeben die obigen Formeln:

$$\hat{\sigma}_p^2 = \frac{2}{9} - \frac{100 \cdot 0,54}{3 \cdot 99} = 0,040404$$

$$a = \frac{2 - 9 \cdot 0,040404}{27 \cdot 0,040404} = 1,5$$

$$\hat{p}_1 = \frac{60 + 1,5}{100 + 4,5} = 0,589$$

---

1) Diese Schätzwerte der Varianz können negative Werte annehmen. Zu diesem Problem der negativen Schätzwerte von Varianzen s. Thionet, P. (1965): S.51-60.

2) Hendricks, Walter A. (1964): S.229.

$$\hat{p}_2 = \frac{30 + 1,5}{100 + 4,5} = 0,301$$

$$\hat{p}_3 = \frac{10 + 1,5}{100 + 4,5} = 0,110$$

Die Berechnung zeigt, daß die Schätzwerte von großen p_h-Werten verringert werden, während kleine vergrößert werden.

Weitere Beispiele - aus dem Gebiet der Agrarstatistik -, in denen die Betaverteilung angewandt wird, werden in den Abschnitten 3.10.1.2.2 und 3.10.2 angeführt.

Um die Größe des Bias zu finden, bilden wir den Erwartungswert:

$$E\hat{p}_h = \frac{1}{n+La} E(n_h+a) = \frac{nP_h+a}{n+La}$$

Als Ausdruck für den Bias über alle L Schichten erhalten wir gemäß dessen Definition:

$$B_{(\hat{p})} = E\sum_h^L (\hat{p}_h - P_h) = \Sigma(E\hat{p}_h - P_h) = \Sigma\left[\frac{nP_h+a}{n+La} - P_h\right]$$

$$= \sum_h^L \left[\frac{nP_h+a - (nP_h+LaP_h)}{n+La}\right] = \sum_h^L \frac{a(1-L\cdot P_h)}{n+La} = 0$$

Der Erwartungswert wird 0, somit ist die Schätzung unverzerrt.

Wir haben also eine Möglichkeit gefunden, die A-posteriori-Schichtung mit all ihren Vorteilen (geringe Kosten, geringer Zeitaufwand, große Wirksamkeit) anzuwenden, auch wenn die zusätzliche Information über die Schichtanteilssätze nicht von vornherein gegeben ist, sondern diese aus den Stichprobendaten geschätzt werden muß.

Diese Methode des Schätzens der Anteilssätze aus Stichprobendaten mit Hilfe von Verteilungsmodellen kann noch verfeinert und auf Unterschichten (Unterkategorien) ausgedehnt werden[1]. Diese weitere Unterteilung der Schichten kommt wohl weniger in der Wirtschafts- und Sozialstatistik als in der naturwissenschaftlichen Statistik und/oder bei Gruppierungsverfahren (s. Abschnitt 3.3.3) vor.

---

1) G. Blumenthal, Soul (1968): S.543ff.

3.7 A-posteriori-Schichtung bei Auswahl mit ungleichen Wahrscheinlichkeiten

3.7.1 Problemstellung

Bisher haben wir die A-posteriori-Schichtung nur bei einer Auswahl mit gleicher Auswahlwahrscheinlichkeit der Merkmale betrachtet. Die Theorie dieser Stichprobenmethode ist in der Literatur umfassend dargestellt.

Sind nun die Auswahlelemente verschieden groß und will oder kann man sie nicht in gleich große Elemente aufspalten, so spricht vieles dafür, sie auch mit den ihnen gemäßen - also ungleichen Wahrscheinlichkeiten - auszuwählen. Die Ausgangschancen der verschiedenen Elemente in der Grundgesamtheit sind ungleich, aber sie müssen bekannt und von 0 verschieden sein. Wenn diese Bedingungen nicht erfüllt sind, befinden wir uns nicht mehr auf dem Gebiet der Zufallsstichproben[1]. Das gebräuchlichste Verfahren bei dieser Auswahl mit ungleichen Wahrscheinlichkeiten ist die Auswahl mit der Wahrscheinlichkeit proportional zur Größe der Elemente (probability proportional to size). Dieses Verfahren wollen wir im wesentlichen hier behandeln. Historisch gesehen hat das "PPS-Verfahren" dem Auswahlverfahren mit ungleichen Wahrscheinlichkeiten auch den Namen geliefert[2].

Ist die Auswahlwahrscheinlichkeit der untersuchten Variablen nicht bekannt, so kann man bei vorsichtiger Abwägung der Umstände die Wahrscheinlichkeitswerte einer anderen Variablen nehmen, wenn diese mit der untersuchten Variablen eng korreliert ist[3].

Dieser Gedankengang wurde nun zur Methode ausgebaut. Man fand, daß man mit ihr die Variabilität der Merkmale kontrollieren, sprich einschränken, kann und so eine größere Genauigkeit der Schätzwerte erzielt.[4] Im Prinzip ist es derselbe Gedankengang, der zur Bildung von ungleichen Schichten, d.h. zur Schichtung

---

1) Vgl. Désabie, J. (1966): S.226.

2) Fellegi, Ivan P. (1963): S.193.

3) Vgl. Sukhatme, Pandurang V. (1963): S.64.

4) Vgl. Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a): S.341.

mit ungleichen Gewichten geführt hat[1].

Verbindungen in der gedanklichen Konzeption bestehen auch zwischen diesem Auswahlverfahren und der Verhältnisschätzung, da die neuen Variablen der Form y_i/z_i den Schätzwerten der Verhältnisschätzung entsprechen (s. Kapitel 2.3.3 Verhältnisschätzung und hier insbesondere Abschnitt 2.3.3.5).

Wie wir schon angedeutet haben, ist der Grund für die Anwendung dieses Verfahrens die Erhöhung der Genauigkeit der Schätzwerte. Wie empirische Untersuchungen zeigen, ist die Reduktion der Varianz beträchtlich. Am Institut für Wirtschafts- und Sozialstatistik, Tübingen, wurden aus heuristischen Gründen solche Untersuchungen angestellt und zwar am Material der belgischen Viehzählung. Dabei hat sich gezeigt, daß sich die relative Standardabweichung bei einer Auswahl der Merkmale mit ungleicher Wahrscheinlichkeit im Vergleich zu einer mit gleicher Auswahlwahrscheinlichkeit um das Zehnfache verringert[2]. Verbesserungen um das 5- bis 6fache an Genauigkeit zeigen Untersuchungen von Hartley und Rao[3] und Horvitz und Thompson[4]. Alle diese Ergebnisse deuten auf die beträchtliche Wirksamkeit dieses Verfahrens hin.

Die Auswahltechnik bei der Auswahl proportional zur Größe der Elemente (pps) kann verschieden ausgeführt werden. Das wohl am meisten angewandte Verfahren ist folgendes: die N Einheiten der Grundgesamtheit werden in eine Liste eingetragen. Die Liste muß einer Zufallsordnung folgen. Es darf auf jeden Fall keine Ordnung der Merkmale vorliegen, die mit der Auswahlwahrscheinlichkeit korreliert ist, da man sonst Gefahr läuft, einen Bias einzuführen[5]. Die Merkmalswerte werden aufsummiert, und bei den

---

1) Vgl. Thionet, M.P. (1953a): S.246.

2) Vgl. auch Zarkovich, S.S. (1960): S.57, der einen theoretischen Wirksamkeitsvergleich der bekanntesten Stichprobenmethoden durchgeführt hat.

3) Hartley, H.O./Rao, J.N.K. (1959): S.58.

4) Horvitz, D.G./Thompson, D.J. (1952): S.682ff..

5) Es soll nur an das bekannte Beispiel der Kilometersteine erinnert werden. Wenn jeder 10.Stein ausgewählt werden soll, und man mit dem 10.Stein beginnt, fallen in die Stichprobe nur Kilometersteine, jedoch keine Hundertmetersteine. Oder an das Beispiel von Schott, der in Mannheim eine Buchstabenauswahl durchführte und dabei nicht berücksichtigte, daß die Anfangsbuchstaben der Familiennamen nicht zufällig verteilt waren, s. Schott, S. (1923): S.44.

kumulierten Werten wird dann systematisch ausgewählt, zum Beispiel mit Hilfe von Schrittziffern. Dabei haben größere Merkmalsbeträge eine größere Chance in die Auswahl zu gelangen als kleinere. Ein Beispiel zeigt das sofort:
Die Merkmalswerte seien

26,3,8,5,17,11,6,5,9,11,41,2,7,....

Die Schrittziffer sei 20, der Zufallsstart sei 4. Dann lautet die kumulierte Reihe der Merkmalswerte (Spalte 2) und die Reihe der kumulierten Schrittziffern (Spalte 3), die die in die Stichprobe fallenden Merkmalswerte (Spalte 4) ermittelt:

Merkmalswerte	kumulierte Merkmalswerte	Zufallsstart u. kumulierte Schrittziffern		ausgewählte Merkmale
1	2	3		4
26 ←	26 ←	4		26
		24	20	26
3	29			
8	37		20	
5	42			
17 ←	59 ←	44	20	17
11 ←	70 ←	64		11
6	76		20	
5	81			
9 ←	90 ←	84		9
11	101		20	
41 ←	142 ←	104	20	41
		124	20	41
2 ←	144 ←	144		2
7	151			

Wir sehen also, daß die großen Merkmalswerte eine größere Auswahlchance haben, daß z.B. 26 und 41 zweimal in die Stichprobe fallen.

Die anderen verwendeten Auswahltechniken sind wesentlich unhandlicher und auf den ersten Blick nicht so einleuchtend wie die oben beschriebene[1)].

Insgesamt gesehen sind die Probleme der Auswahl mit ungleichen

---

1) Vgl. Rao, J.N.K. (1963): S.202/203.

Wahrscheinlichkeiten bisher in der Literatur noch nicht vollständig behandelt worden[1]. Kein theoretisches Verfahren ist ganz befriedigend. Oft werden nur die unpraktikablen aber theoretisch gut gelösten Fälle mit n=2 behandelt[2]. Obwohl die Stichproben mit variabler Auswahlwahrscheinlichkeit schon in den zwanziger Jahren entwickelt wurden, werden die Probleme in den Lehrbüchern meist nur kurz abgehandelt und an den Schluß der Betrachtungen gesetzt[3].

Die Probleme sind also keineswegs geklärt, die Methoden stark in der Entwicklung begriffen. Im folgenden wollen wir versuchen, hierzu ein Modell aufzustellen, das auf die A-posteriori-Schichtung zutrifft.

3.7.2 Erwartungswert

In dem Modell[4] gehen wir davon aus, daß die Stichprobe ausgewählt vorliegt und daß die Merkmale mit ungleichen Wahrscheinlichkeiten ausgewählt worden sind. Die Auswahlwahrscheinlichkeiten p_i sind also gegeben.

Die Stichprobe soll nun geschichtet werden, das heißt, die Merkmalswerte mit ihren dazugehörigen Auswahlwahrscheinlichkeiten werden in Schichten aufgeteilt.

---

1) Vgl. Thionet, P. (1967): S.16.

2) So z.B. bei Fellegi, Ivan P. (1963): S.188ff.; s. auch die ausführliche Kritik dieses Auswahlverfahrens bei Sukhatme, P.V. (1966): S.381/382.

3) Vgl. Stuart, Alan (1964): S.773.

4) Das Modell wurde aufgestellt im Zusammenhang mit einer Korrespondenz mit folgenden Professoren: W.G. Cochran, Harvard-Universität; T. Dalenius, Universität Stockholm; M.R. Sampford, Universität Liverpool; P. Thionet, Universität Poitiers; B.V. Sukhatme, Iowa-State-Universität. Bei der Korrespondenz ergab sich, daß die Probleme noch im Fluß sind und keine Lösungen vorliegen. Erschwerend wirkte sich aus, daß die meisten sich mit den Problemen noch nicht ausführlich befaßt hatten und/oder von verschiedenen Ansatzpunkten ausgingen.

Es gilt

$$\sum_{i=1}^{n} p_i = \sum_{h=1}^{L} \sum_{i=1}^{n_h} p_{hi}$$

und

$$p_h = \sum_i p_{hi}$$

Die Summe der Wahrscheinlichkeiten in der Stichprobe soll gleich 1 sein. Demnach müssen die Wahrscheinlichkeiten mit einem Proportionalitätsfaktor multipliziert werden:

$$\sum_h \sum_i c p_{hi} = 1$$

Somit

$$c = \frac{1}{\sum_h \sum_i p_{hi}} = \frac{1}{\sum_h p_h} \qquad (41)$$

Die Wahrscheinlichkeiten, mit denen wir die Merkmale hochrechnen müssen, sind dann:

$$\pi_{hi} = \frac{p_{hi}}{p_h} \sum_h p_h \qquad (42)$$

Führen wir die transformierte Variable ein, wie dies bei dem Modell mit variabler Auswahlwahrscheinlichkeit üblich ist, so wird diese

$$z_{hi} = \frac{y_{hi}}{N_h \pi_{hi}} = \frac{y_{hi}}{N_h p_{hi}} \frac{p_h}{\sum_h p_h} \qquad (43)$$

Der Schätzwert $\bar{y}_{p/pps}$ der A-posteriori-Schichtung bei ungleichen Auswahlwahrscheinlichkeiten (pps im Index = probability proportional to size) wird dann[1]:

$$\bar{y}_{p/pps} = \sum_h \frac{N_h}{N} \bar{z}_h = \sum_h \frac{N_h}{N} \frac{1}{n_h} \sum_i^{n_h} z_{hi}$$

$$= \Sigma \frac{N_h}{N} \frac{1}{n_h} \Sigma \frac{y_{hi}}{N_h \cdot p_{hi}} \frac{p_h}{\Sigma p_h}$$

---

[1]) Vgl. dazu auch die Schätzwerte der A-priori-Schichtung bei Cochran, William G. (1964): S.260 und bei Rao, J.N.K./Hartley, H.O./Cochran, W.G. (1962): S.483ff..

oder

$$\bar{y}_{p/pps} = \frac{1}{N} \sum_{h}^{L} \frac{1}{n_h} \sum_{i}^{n_h} \frac{y_{hi}}{\pi_{hi}} = \frac{1}{N} \Sigma \frac{1}{n_h} \Sigma \frac{y_{hi}}{p_{hi}} \frac{p_h}{\Sigma p_h} \qquad (44)$$

Der Schätzwert für den Totalwert Y ist folglich:

$$\hat{Y}_{p/pps} = \sum_{h}^{L} \frac{1}{n_h} \sum_{i}^{n_h} \frac{y_{hi}}{\pi_{hi}} \qquad (45)$$

Betrachten wir in (44) die $p_h/\Sigma p_h$ als Schichtanteilssätze N_h/N, dann erhalten wir den Schätzwert $\bar{y}'_{p/pps}$, der in der Literatur bei Modellen mit variabler Auswahlwahrscheinlichkeit zuweilen zu finden ist[1]:

$$\bar{y}'_{p/pps} = \sum_{h}^{L} \frac{1}{n_h} \sum_{i}^{n_h} \frac{y_{hi}}{N p_{hi}} \frac{N_h}{N}$$

$$= \Sigma \frac{N_h}{N} \frac{1}{n_h} \Sigma \frac{y_{hi}}{N p_{hi}} \qquad (46)$$

Diesem Schätzwert fehlt die Gewichtung bezüglich der Schichtwahrscheinlichkeiten. Er ist wohl nur anwendbar bei einer A-priori-Schichtung bzw. bei der A-posteriori-Schichtung, wenn eine Gewichtung der Schichtwahrscheinlichkeiten nicht erforderlich ist.

Der Schätzwert $\bar{y}_{p/pps}$ in (44) ist unverzerrt, da der Erwartungswert in den Schichten Y_h ergibt. Über alle Schichten summiert und durch N dividiert, erhalten wir wie folgt den wahren Mittelwert $\bar{Y}$:

$$E(\bar{y}_{p/pps}) = E\left[\frac{1}{N} \sum_{h}^{L} \frac{1}{n_h} \sum_{i}^{n_h} \frac{y_{hi}}{\pi_{hi}}\right] = \frac{1}{N} \sum_{h}^{L} \frac{1}{n_h} \sum_{i}^{n_h} E\left(\frac{y_{hi}}{\pi_{hi}}\right)$$

$$= \frac{1}{N} \sum_{h}^{L} \frac{\pi_{hi} y_{hi}}{\pi_{hi}} = \frac{1}{N} \sum_{h}^{L} Y_h = \frac{Y}{N}$$

$$= \bar{Y}$$

---

[1] So etwa bei den Autoren Horvitz, D.G./Thompson, D.J. (1952): S.668 mit ihrer Schätzfunktion T_3, obwohl sie nicht von der A-posteriori-Schichtung ausgehen und somit ihre Betrachtung unter anderen Voraussetzungen anstellen.

3.7.3 Varianz und Schätzwert der Varianz

Der Ansatz zur Berechnung der Varianz ist derselbe wie bei der A-posteriori-Schichtung mit einfacher Zufallsauswahl (gleiche Wahrscheinlichkeiten), nämlich

$$V(\bar{y}_{p/pps}) = E\left[\sum_h \frac{N_h^2}{N^2} \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}\right] \qquad (5)$$

Da die n_h bei der A-posteriori-Schichtung von vornherein nicht bekannt sind, brauchen wir $E(\frac{1}{n_h})$, und wir erhalten dann als Näherungswert für die Varianz (s. Abschnitt 3.4.3 Varianz):

$$V(\bar{y}_{p/pps}) = \frac{N-n}{N \cdot n}\left[\sum_h \frac{N_h}{N} S_h^2 + \frac{1}{n} \sum_h \frac{N-N_h}{N} S_h^2\right] \qquad (47)$$

Dabei gilt jetzt für[1]:

$$S_h^2 = \sum_i^{N_h} \pi_{hi}(z_{hi} - \bar{Z}_h)^2 \qquad (48)$$

Für $\bar{Z}_h$ können wir auch schreiben:

$$\bar{Z}_h = E(z_{hi}) = \sum_i^{N_h} \pi_{hi} \frac{y_{hi}}{N_h \pi_{hi}} = \bar{Y}_h$$

somit

$$S_h^2 = \sum_i^{N_h} \pi_{hi}(z_{hi} - \bar{Y}_h)^2 \qquad (48a)$$

Der Schätzwert für S_h^2 ist

$$s_h^2 = \frac{1}{n_h - 1} \sum_i^{n_h} (z_{hi} - \bar{z}_h)^2 \qquad (49)$$

bzw.

$$s_h^2 = \frac{1}{n_h - 1} \sum_i^{n_h} (z_{hi} - \bar{y}_h)^2 \qquad (49a)$$

Der Schätzwert der Varianz bei ungleichen Auswahlwahrscheinlichkeiten wird also, indem wir (49) einsetzen:

---

[1]) Vgl. im folgenden Sukhatme, Pandurang V. (1963): S.128/129.

$$v(\bar{y}_{p/pps}) = \frac{N-n}{N \cdot n}\left[\sum_h \frac{N_h}{N} s_h^2 + \frac{1}{n}\sum_h \frac{N-N_h}{N} s_h^2\right] \qquad (50)$$

bzw. beim Totalwert:

$$v(\hat{Y}_{p/pps}) = \frac{N-n}{n}\left[\sum_h N_h s_h^2 + \frac{1}{n}\sum_h (N-N_h) s_h^2\right] \qquad (51)$$

Bei der Behandlung der Varianz kann der Korrekturfaktor (N-n)/N auch vernachlässigt werden, da wir es
a) mit einem Näherungswert zu tun haben, und
b) wir S_h^2 für den Fall mit Zurücklegen genommen haben[1].

In Abschnitt 3.10.2 werden wir das eben behandelte Modell an einem Beispiel aus der Agrarstatistik demonstrieren und seine Wirksamkeit zeigen.

3.8 Kombination der A-posteriori-Schichtung mit anderen Hochrechnungsverfahren

Die A-posteriori-Schichtung läßt sich - und das ist einer ihrer besagten Vorteile - mit den anderen bekannten Hochrechnungsarten (Kapitel 2.3) zu kombinierten Modellen verwenden. Diese Möglichkeit ist bei den anderen Hochrechnungsmethoden nicht in der Weise gegeben, teilweise weil der Formel- wie der Rechenaufwand im Verhältnis zum Genauigkeitsgewinn unangemessen steigt, teilweise erscheint es "doppelt genäht", neben einer Regression der Variablen noch ihr Verhältnis oder ihre Differenz zu bilden.

Allgemein kann gesagt werden, daß der Zuwachs an Wirksamkeit solcher kombinierter Verfahren unterschiedlich sein wird. So wird sich eine Kombination mit der Regressionsschätzung nur bei großem Schichtungseffekt lohnen. Für eine solche Kombination der A-posteriori-Schichtung mit anderen Hochrechnungsverfahren

---

[1]) Vgl. dazu Durbin, J. (1953): S.266f..

spricht vor allem ihr geringer Arbeitsaufwand. Die folgenden Formeln sind analog hergeleitet; dies ist wegen der algebraischen Form der Formeln möglich.[1] Bei den einzelnen Modellen benützen wir immer ein kombiniertes geschichtetes Hochrechnungsverfahren[2], da die Schichtbesetzung bei der A-posteriori-Schichtung von vornherein nicht bekannt ist, und bei geringer Schichtbesetzung das kombinierte Verfahren zweckmäßiger ist als das separate.

Unser früheres Schema (Abb.4) erweitert sich dann zu:

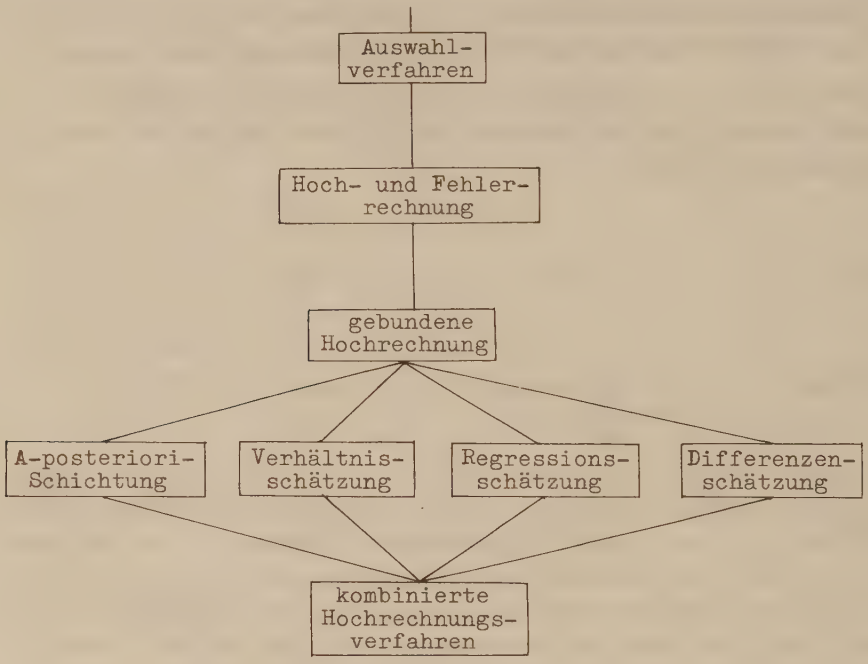


Abb.13 Erweiterung des Stichprobenmodells durch kombinierte Hochrechnungsverfahren

Bei den folgenden kurzen Wirksamkeitsbetrachtungen sei auf Abschnitt 3.5 Wirksamkeit hingewiesen; die Ausführungen hier sollen als Ergänzung dazu verstanden werden.

---

[1] Die Variable x bedeutet immer die zusätzliche Information; wir beziehen uns in der Notation auf Kapitel 2.3; ferner gilt $S_h^2 = S_{yh}^2$.

[2] Vgl. die Abschnitte 2.3.3.6 und 2.3.4.5.

3.8.1 A-posteriori-geschichtete Verhältnisschätzung

Die auf den ersten Blick einleuchtendste Kombination ist die A-posteriori-Schichtung mit der Verhältnisschätzung. Dieses Verfahren wird in der Praxis zuweilen auch angewandt[1]. Man bildet in jeder Schicht das Verhältnis der untersuchten Variablen y mit der ihr korrelierten Variablen x der zusätzlichen Information (s. Abschnitt 2.3.3 Verhältnisschätzung).

Zwischen beiden Verfahren besteht eine große Analogie. Bei der Verhältnisschätzung wird normalerweise ein quantitatives Merkmal als zusätzliche Information genommen; nimmt man dazu ein qualitatives Merkmal, so erhalten wir im Prinzip eine Schichtung. y_i/x_i geht dann über in $(M_i/M)\cdot y_i$, wobei M_i/M den Anteilssatz darstellt[2].

Die Formeln dieses Verfahrens lauten:

für den Schätzwert des Mittelwertes

$$\bar{y}_{pv} = \frac{\sum_h^L \frac{N_h}{N} \bar{y}_h}{\sum_h^L \frac{N_h}{N} \bar{x}_h} \bar{X} = \frac{\sum_h^L N_h \bar{y}_h}{\sum_h^L N_h \bar{x}_h} \bar{X} \qquad (52)$$

für die Varianz des Mittelwertes

$$V(\bar{y}_{pv}) = \frac{N-n}{N \cdot n} \left[\sum_h^L \frac{N_h}{N} (S_{yh}^2 + F^2 S_{xh}^2 - 2FS_{xyh}) + \frac{1}{n} \sum_h^L \frac{N-N_h}{N} (S_{yh}^2 + F^2 S_{xh}^2 - 2FS_{xyh}) \right] \qquad (53)$$

Dabei ist $F = \bar{Y}/\bar{X}$.

Über die Wirksamkeit der a-posteriori-geschichteten Verhältnisschätzung im Vergleich zur einfachen Verhältnisschätzung läßt sich theoretisch nur feststellen, daß, je größer der Schichtungseffekt ist, d.h., je mehr die Gewichte N_h/N zur Verringerung der Varianz beitragen, um so wirksamer wird das kombinier-

---

1) S. Thionet, M.P. (1953b): S.102ff..

2) Vgl. Williams, W.H. (1962): S.624ff. und Thionet, M.P. (1953a): S.246.

te Verfahren sein[1]. Hier bestimmen ausschließlich die Gewichte über die Wirksamkeit und daher Zweckmäßigkeit des a-posteriori-geschichteten Hochrechnungsverfahrens.

Beim Vergleich der Wirksamkeit der A-posteriori-Schichtung mit der a-posteriori-geschichteten Verhältnisschätzung gilt wieder

$$V(\bar{y}_p) > V(\bar{y}_{pv})$$

$$\left[\Sigma \frac{N_h}{N} S^2_{yh} + \frac{1}{n}\Sigma \frac{N-N_h}{N} S^2_{yh}\right] > \left[\Sigma \frac{N_h}{N}(S^2_{yh} + F^2 S^2_{xh} - 2FS_{xyh}) + \frac{1}{n}\Sigma \frac{N-N_h}{N}(S^2_{yh} + F^2 S^2_{xh} - 2FS_{xyh})\right]$$

Nach einigen Umformungen erhalten wir:

$$\left[\Sigma \frac{N_h}{N} S_{yxh} + \frac{1}{n}\Sigma \frac{N-N_h}{N} S_{xyh}\right] > \frac{F}{2}\left[\Sigma \frac{N_h}{N} S^2_{xh} + \frac{1}{n}\Sigma \frac{N-N_h}{N} S^2_{xh}\right]$$

Unter der Annahme eines von der Schichtung unabhängigen Korrelationskoeffizienten zwischen x und y, gilt:

$$\varrho\left[\Sigma \frac{N_h}{N} S_{xh}S_{yh} + \frac{1}{n}\Sigma \frac{N-N_h}{N} S_{xh}S_{yh}\right] > \frac{F}{2}\left[\Sigma \frac{N_h}{N} S^2_{xh} + \frac{1}{n}\Sigma \frac{N-N_h}{N} S^2_{xh}\right]$$

oder

$$\varrho > \frac{F\left[\Sigma \frac{N_h}{N} S^2_{xh} + \frac{1}{n}\Sigma \frac{N-N_h}{N} S^2_{xh}\right]}{2\left[\Sigma \frac{N_h}{N} S_{xh}S_{yh} + \frac{1}{n}\Sigma \frac{N-N_h}{N} S_{xh}S_{yh}\right]} \qquad (54)$$

Bei Vernachlässigung des Summengliedes mit 1/n in Zähler und Nenner erhalten wir einen analogen Ausdruck für die Wirksamkeit, der uns von der einfachen Verhältnisschätzung her bekannt ist:

$$\varrho > \frac{F\ \Sigma N_h S^2_{xh}}{2\ \Sigma N_h S_{xh} S_{yh}} \qquad (54a)$$

---

[1]) Vgl. Abschnitt 3.5 Wirksamkeit, wo wir die gleiche Feststellung trafen bezüglich der A-posteriori-Schichtung und der einfachen Verhältnisschätzung.

Die a-posteriori-geschichtete Verhältnisschätzung ist also wirksamer als die A-posteriori-Schichtung, wenn der Korrelationskoeffizient ϱ größer ist als F/2 des Verhältnisses der gewichteten Varianz der Variablen x (zusätzliche Information) zu dem Produkt der gewichteten Einzelvarianzen von x und y.

3.8.2 A-posteriori-geschichtete Differenzenschätzung

Wie wir schon an früherer Stelle ausgeführt haben (Abschnitt 2.3.5 Differenzenschätzung), sind Differenzenschätzung und Verhältnisschätzung verwandte Verfahren. Die Bemerkungen des vorhergehenden Abschnittes gelten daher zum großen Teil auch für dieses Verfahren. Der Grundgedanke ist hier der, daß man in den einzelnen Schichten Differenzen bildet und diese gewichtet hochrechnet.

Die Berechnungsformeln lauten:

für den Mittelwert:

$$\bar{y}_{pd} = \sum_h^L \frac{N_h}{N} \bar{y}_h + k(\bar{X} - \sum_h^L \frac{N_h}{N} \bar{x}_h) \qquad (55)$$

für die Varianz des Mittelwertes

$$V(\bar{y}_{pd}) = \frac{N-n}{N \cdot n}\left[\sum_h^L \frac{N_h}{N}(S_{yh}^2 + k^2 S_{xh}^2 - 2kS_{xyh}) + \frac{1}{n}\sum_h^L \frac{N-N_h}{N}(S_{yh}^2 + k^2 S_{xh}^2 - 2kS_{xyh})\right] \qquad (56)$$

k ist wie üblich bei der Differenzenschätzung eine Konstante.

Bei der Wirksamkeit der a-posteriori-geschichteten Differenzenschätzung im Vergleich mit der einfachen Differenzenschätzung kommt es wieder auf die Wirkung der Gewichte N_h/N an.

Für den Vergleich der Wirksamkeit der A-posteriori-Schichtung mit der a-posteriori-geschichteten Differenzenschätzung gilt analog zur Verhältnisschätzung (54a) als Gradmesser die Näherungsformel (bei Vernachlässigung des Summengliedes mit 1/n):

$$\varrho > \frac{k\ \Sigma N_h S_{xh}^2}{2\ \Sigma N_h S_{xh} S_{yh}} \qquad (57)$$

d.h., es kommt wieder auf das Verhältnis der gewichteten Varianzen der Variablen x und y an.

3.8.3 A-posteriori-geschichtete Regressionsschätzung

Die Ausführungen über den Aufwand an Rechenarbeit bei der Regressionsschätzung sind bekannt. In verstärktem Maße gilt das für die a-posteriori-geschichtete Regressionsschätzung. Bei zweckmäßiger Anwendung haben wir es jedoch mit dem wirksamsten der behandelten Verfahren zu tun.[1] Bei der Schätzung wollen wir den Regressionskoeffizienten b für alle Schichten als gleich annehmen, da bei geringer Schichtbesetzung ein Bias auftreten kann.

Die Formeln für den Mittelwert und die Varianz lauten:

für den geschätzten Mittelwert

$$\bar{y}_{pr} = \sum_h \frac{N_h}{N} \bar{y}_h + b\left(\bar{X} - \sum_h \frac{N_h}{N} \bar{x}_h\right) \qquad (58)$$

für den Näherungswert der Varianz, wenn wir den Korrelationskoeffizienten ϱ für alle Schichten als gleich annehmen:

$$V(\bar{y}_{pr}) = \frac{N-n}{N \cdot n}\left[\sum_h \frac{N_h}{N} S_{yh}^2(1-\varrho^2) + \frac{1}{n} \sum_h \frac{N-N_h}{N} S_{yh}^2(1-\varrho^2)\right]$$

$$= \frac{N-n}{N \cdot n}(1-\varrho^2)\left[\Sigma \frac{N_h}{N} S_{yh}^2 + \frac{1}{n} \Sigma \frac{N-N_h}{N} S_{yh}^2\right] \qquad (59)$$

Für den Vergleich der Wirksamkeit der a-posteriori-geschichteten Regressionsschätzung mit der einfachen Regressionsschätzung gilt wieder, daß erstere um so wirksamer ist, je mehr die Gewichte N_h/N durchschlagen. Sie hängt also wieder von der Größe des Schichtungseffekts ab.

---

[1]) Ein ähnliches Verfahren - mit anderen Gewichten - ist zu finden bei Sukhatme, Pandurang V. (1963): S.204ff..

Vergleichen wir die A-posteriori-Schichtung mit der a-posteriori-geschichteten Regressionsschätzung, so sehen wir sofort, daß letztere in jedem Falle wirksamer ist, solange $\varrho > 0$, denn es gilt

$$V(\bar{y}_p) > V(\bar{y}_{pr})$$

$$\left[\Sigma \frac{N_h}{N} s^2_{yh} + \frac{1}{n} \Sigma \frac{N-N_h}{N} s^2_{yh}\right] > (1-\varrho^2)\left[\Sigma \frac{N_h}{N} s^2_{yh} + \frac{1}{n} \Sigma \frac{N-N_h}{N} s^2_{yh}\right]$$

3.9 Anwendung

Bisher haben wir die theoretischen Grundlagen geschaffen, mit Hilfe derer die A-posteriori-Schichtung angewandt werden kann. Nun sind in der offiziellen Statistik wenig Beispiele zu finden, bei denen die A-posteriori-Schichtung als Hochrechnungsverfahren tatsächlich angewandt wird. So kann man die Fälle aus der praktischen Stichprobenerhebung nicht herausschälen, in denen es geschickt und wegen der höheren Wirksamkeit angebracht wäre, dieses Verfahren für bestimmte Kategorien von Stichprobenuntersuchungen zu empfehlen.

Eine Anfrage an Fachleute aus verschiedenen Ländern erbrachte hierüber ebenfalls keine Information[1]. Es wurde zwar angedeutet, daß dieses Verfahren angewandt werde, aber konkrete Veröffentlichungen über dessen Anwendung bei praktischen Stichprobenerhebungen wurden nicht genannt.

In der amtlichen deutschen Statistik[2] scheint die Anwendung der A-posteriori-Schichtung (Gruppierung nach der Auswahl) bei laufenden Verfahren auf den Mikrozensus beschränkt zu sein, wenn man bei bestimmten Verfahren von der Schichtung nach Län-

---

1) Um Informationen darüber zu erhalten, wurden angeschrieben die Professoren Cochran, Sedransk (USA); Thionet, Désabie (Frankreich); Sukhatme (FAO, Rom); Dalenius (Schweden); Sampford (Großbritannien).

2) Statistisches Bundesamt (Hrsg.)(1960): S.48 und S.377.

dern einmal absieht. Ferner wurde die A-posteriori-Schichtung bei der Repräsentativerhebung über den Güterverkehr mit Kraftfahrzeugen im Juli 1952 angewandt; dabei wurde auf die guten Ergebnisse hingewiesen, die man mit Hilfe dieses Verfahrens erzielt hat.[1)]

In der theoretischen Literatur ist über die Anwendung ebenfalls wenig zu finden. Ab und zu wird ein Beispiel angeführt, aus dem aber nicht klar hervorgeht, ob es als Demonstrationsbeispiel gebildet ist oder ob es aus der offiziellen Statistik stammt, da dieses Beispiel durch alle Verfahren hindurch gezogen wird.[2)] In der theoretischen Literatur gilt oft was Thionet im Hinblick auf einen Autor schreibt, dessen Lektüre er wenig angenehm empfand: "des calculs, mais peu d'idees".[3)]

Eine zwangsläufige Anwendung der A-posteriori-Schichtung ist wohl beim Non-response-Problem gegeben, da die beiden Schichten der Antworter bzw. Nichtantworter erst nach der Auswahl bekannt werden (s. Abschnitt 3.6.2 Schätztheorie der Schichtung).[4)]

In der amtlichen deutschen Statistik bieten sich eine Reihe von Anwendungsmöglichkeiten der A-posteriori-Schichtung an. Bei fast allen Stichprobenerhebungen, die nicht schon gebunden hochgerechnet werden, ist es prinzipiell möglich, nachträglich Schichten zu bilden. Wie das Arbeitsgebiet der Bundesstatistik[5)] zeigt, wäre dies bei einer Reihe von Stichprobenerhebungen möglich. Dabei wollen wir von einer Schichtung nach Bundesländern absehen, da die Landesämter die Erhebungen meist durchführen, und deshalb dieses Schichtungsmerkmal in jeder Untersuchung vorhanden ist. Da die Tendenz der amtlichen Statistik dahin geht,

---

1) Statistisches Bundesamt (1954): S.3ff..

2) So z.B. bei Yates, Frank (1965): S.152 und S.205.

3) Thionet schrieb dies in einem Brief vom 21.3.69 an den Verfasser im Rahmen einer Korrespondenz über das Modell mit ungleichen Auswahlwahrscheinlichkeiten (vgl. Abschnitt 3.7.2).

4) Die A-posteriori-Schichtung kann allgemein angewandt werden auf dem Gebiet der Fehlerkorrektur, ob dies bei der Nachprüfung von Angabefehlern ist (s. Abschnitt 3.10.2) oder bei der "automatischen Entdeckung und Korrektur" von Fehlern, die bei Untersuchungen auftreten (vgl. Szameitat, K./Zindler, H.J. (1966): S.395ff.).

5) Statistisches Bundesamt (Hrsg.)(1966): S.310ff..

immer mehr Stichprobenerhebungen durchzuführen, teils um zusätzliche und/oder aktuellere Ergebnisse zu erhalten, teils um Vollerhebungen zu ersetzen, und ferner Teilerhebungen auf Stichprobenerhebungen umgestellt werden sollen[1], wird wohl in der Zukunft auch auf verfeinerte Hochrechnungsmethoden größeren Wert gelegt werden.

Ein weites und wichtiges Anwendungsgebiet der A-posteriori-Schichtung werden unserer Ansicht nach die Stichprobenmodelle mit variabler Auswahlwahrscheinlichkeit werden. Bei variabler Auswahlwahrscheinlichkeit ist die Schichtung nach den untersuchten Variablen y_i oft nicht effektiv. Denn es ist möglich, daß die mit den Kehrwerten der Wahrscheinlichkeiten multiplizierten Merkmale den konzipierten Schichtgrößen nicht mehr entsprechen. Daher ist es vorteilhaft, die transformierten Variablen z_i zu schichten, wobei

$$z_i = \frac{y_i}{Np_i}$$

diese transformierte Variable darstellt (vgl. (43) in Abschnitt 3.7.2), die erst nach erfolgter Auswahl (a-posteriori) berechnet werden kann. Durch die Schichtenbildung kann der Schichtungseffekt mit Hilfe der neuen Variablen vergrößert werden. Gerade bei zweistufigen Verfahren, bei denen die Primäreinheiten mit variabler Wahrscheinlichkeit ausgewählt wurden und diese oft einen heterogenen Charakter aufweisen, bringt eine solche Schichtung der Primäreinheiten eine starke Reduzierung der Varianz[2]. Dabei ist es für eine Anwendung der A-posteriori-Schichtung günstig, wenn die Schichten mit 25-50 Einheiten besetzt sind, denn dann erhalten wir einen Genauigkeitsgewinn ähnlich dem der A-priori-Schichtung[3]. Hält man diese Richtzahl fest, kann man danach die Anzahl der Schichten bestimmen.

Im folgenden Abschnitt wollen wir einige Beispiele bringen, bei denen die hergeleiteten Formeln angewandt werden, und die Wirksamkeit dieser Schätzmethode demonstriert wird.

---

1) Statistisches Bundesamt (Hrsg.)(1966): S.34.

2) Vgl. Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a): S. 210.

3) Vgl. ebenda S.232.

3.10. Anwendungsbeispiele

Am Institut für Wirtschafts- und Sozialstatistik, Tübingen, bei Professor Strecker wurden gründliche Untersuchungen über Angabefehler bei Viehzählungen in Belgien angestellt.[1] Angabefehler sind die Fehler, die Betriebsinhaber bei der Deklaration ihres Viehbestandes begehen, sei es bewußt oder unbewußt. Der Schwerpunkt der Untersuchungen lag dabei auf methodischem Gebiet. Man wollte Methoden gewinnen, die aktuell und mit geringen Kosten wirksame Schätzwerte liefern. Dieses Material aus der amtlichen belgischen Statistik bietet sich für Demonstrationsbeispiele an, und man ist nicht genötigt, sich auf konstruierte Beispiele zurückziehen zu müssen.

Im folgenden wollen wir mit den bisher erarbeiteten theoretischen Grundlagen a) den Viehbestand (Abschnitt 3.10.1) und b) den Angabefehler (Abschnitt 3.10.2) schätzen und zwar im Fall a) mit gleicher und in b) mit variabler Auswahlwahrscheinlichkeit.

3.10.1 Schätzung des Viehbestandes

In unserem Beispiel befassen wir uns nur mit dem Schweinebestand. Die Auswahleinheiten sind die Anzahl der Schweine in der Gemeinde. Als Auswahlmodell wird die einfache Zufallsauswahl genommen, d.h., jede Gemeinde hat die gleiche Auswahlchance. Der Stichprobenumfang beträgt 150 Gemeinden. Wir behandeln ein einstufiges Modell.

Es ist:

y_i = Schweinebestand in der Gemeinde
M_i = Zahl der Schweinehalter in der Gemeinde
n = 150 = Stichprobenumfang
N = 2 548 = Zahl der Gemeinden in Belgien mit Schweinebestand am 15.Mai 1968
Y = 2 479 025 = Gesamtschweinebestand in Belgien laut Zählung vom 15.Mai 1968
M = 90 012 = Gesamtzahl der Schweinehalter in Belgien laut Zählung vom 15.Mai 1968

---

[1] S. Strecker, H./Steylaerts, R. (1966).

3.10.1.1 Einfache Hochrechnung

Die 150 Gemeinden wurden mit Hilfe der einfachen Zufallsauswahl ausgewählt. Daraus wurde der mittlere Schweinebestand pro Gemeinde, sowie dessen Varianz bzw. die Standardabweichung berechnet. Die Hochrechnung erfolgt, indem der Mittelwert mit N bzw. die Varianz mit N^2 multipliziert wird. Der Schätzwert des Gesamtschweinebestandes für Belgien wird dann:

$$\hat{Y} = N\bar{y} = 2548 \cdot 1277{,}1467$$

$$= 3\ 254\ 170 \text{ Schweine}$$

Dessen Standardabweichung ist:

$$\sqrt{v(\hat{Y})} = s_{\hat{Y}} = 532\ 525 \text{ Schweine}$$

3.10.1.2 A-posteriori-Schichtung

Das Modell der einfachen Zufallsauswahl wird beibehalten. Jedoch werden jetzt die Stichprobengemeinden nach der Auswahl der Größe ihrer Schweinebestände nach in 4 Schichten aufgeteilt. 4 Schichten wurden deshalb gebildet, damit die einzelnen Schichten genügend besetzt sind (vgl. Abschnitt 3.9 Anwendung). In jede Schicht sollten ungefähr 30 bzw. mehr Stichprobeneinheiten fallen. Darüber hinaus wurden die Schichten willkürlich gebildet, da es auf den Vergleich der Verfahren und nicht auf den größtmöglichen Schichtungseffekt ankam. Die Schichtung wird des Vergleichs wegen in den behandelten Fällen in diesem Abschnitt 3.10.1.2 nicht verändert.

Die Ergebnisse für Schichtmittelwert $\bar{y}_h$, Schätzwert der Varianz s_h^2 und Standardabweichung s_h sind in Übersicht 1 zusammengestellt.

Wie wir schon ausführlich dargestellt haben, besteht das Hauptproblem bei der A-posteriori-Schichtung darin, die Schichtanteilssätze als zusätzliche Informationen zu bekommen. Dazu wollen wir 3 Fälle behandeln:

- Hochrechnung mit den "wahren" N_h-Werten

- Hochrechnung mit geschätzten $\hat{N}_h$-Werten
- Hochrechnung mit M_h-Werten einer korrelierten Variablen

Übersicht 1 Berechnete Schichtwerte

Schicht[1)]	Schweinebestand von... bis...	Stichprobengemeinden n_h	Mittelwert $\bar{y}_h$	Varianz s_h^2	Standardabweichung s_h
1	1-150	48	55,9	1990,0	44,6
2	151-400	29	259,8	4554,6	67,5
3	401-1000	34	624,5	31600,2	177,8
4	1001 und mehr	39	3712,0	9549723,3	3090,3
-	-	150	-	-	-

[1)] Der Kürze halber wollen wir bei den folgenden Übersichten nur Schicht 1,2,3,4 schreiben.

3.10.1.2.1 Hochrechnung mit den "wahren" N_h-Werten

Die Stichprobengemeinden wurden schon in 4 Schichten aufgeteilt. Die "wahren" Schichtanteilssätze der Grundgesamtheit können nun bestimmt werden, indem alle Gemeinden Belgiens nach den gleichen Schichtungsmerkmalen sortiert werden. Überlegungen über eine mögliche Fluktuation der Merkmale zwischen den Schichten, die bei einer Bildung der Schichten nach altem Material auftreten, fallen hier weg, da sozusagen am gleichen (neuen) Material geschichtet wird (vgl. dazu Abschnitt 3.2 bei der Behandlung der Vorteile der A-posteriori-Schichtung). Nebenbei sei bemerkt, daß beim Schweinebestand eine starke Fluktuation besteht (vgl. Abschnitt 3.4.5).

Sortieren wir alle Gemeinden Belgiens nach den Schichtmerkmalen der Stichprobe, erhalten wir für die "wahren" N_h-Werte (Übersicht 2):

Übersicht 2 "Wahre" N_h-Werte

Schicht	Stichproben-gemeinden n_h	Zahl der Gemeinden in der Grundgesamtheit N_h
1	48	743
2	29	607
3	34	590
4	39	608
-	150	2548

Berechnen wir den Schweinebestand nach der Schätzformel für die A-posteriori-Schichtung (2), erhalten wir für den Gesamtschweinebestand:

$$\hat{Y}_{p_1} = \sum_h N_h \bar{y}_h = 2\ 824\ 571 \text{ Schweine}$$

Die Standardabweichung (18) wird:

$$\sqrt{v(\hat{Y}_{p_1})} = s_{\hat{Y}_{p_1}} = 308\ 497 \text{ Schweine}$$

3.10.1.2.2 Hochrechnung mit geschätzten $\hat{N}_h$-Werten

Es kann nun der Fall eintreten, daß wir die Information über die Anteilssätze aus der Grundgesamtheit (wie dies im vorhergehenden Abschnitt der Fall war) nicht zur Verfügung stehen haben. In Abschnitt 3.6.2.2 haben wir ein Verfahren beschrieben, auf Grund dessen die Anteilssätze aus der Stichprobe geschätzt werden können.

Wir nehmen als Grundverteilung eine Betaverteilung an und berechnen (s. Abschnitt 3.6.2.2) mit Hilfe von (40) und (34) die Parameter

$$\hat{\sigma}_p^2 = 0{,}0009452$$

und

$$a = 49{,}3432$$

Daraus berechnen wir (35) die $\hat{p}_h$-Werte:

$$\hat{p}_1 = 0{,}280$$

$$\hat{p}_2 = 0{,}226$$

$$\hat{p}_3 = 0{,}240$$

$$\hat{p}_4 = 0{,}254$$

Indem wir diese $\hat{p}_h$-Werte mit der Gesamtzahl der Gemeinden Belgiens multiplizieren, erhalten wir die geschätzten $\hat{N}_h$-Werte (Übersicht 3):

Übersicht 3 Geschätzte $\hat{N}_h$-Werte

Schicht	Stichproben-gemeinden n_h	Zahl der Gemeinden in der Grundgesamtheit (geschätzt) $\hat{N}_h$
1	48	713
2	29	576
3	34	612
4	39	647
-	150	2548

Berechnen wir den Schweinebestand mit den geschätzten Anteilssätzen, erhalten wir für den Gesamtschweinebestand

$$\hat{Y}_{p_2} = \sum_h \hat{N}_h \bar{y}_h = 2\ 973\ 346 \text{ Schweine}$$

und für die Standardabweichung

$$\sqrt{v(\hat{Y}_{p_2})} = s_{\hat{Y}_2} = 317\ 951 \text{ Schweine}$$

3.10.1.2.3 Hochrechnung mit den M_h-Werten einer korrelierten Variablen

An früherer Stelle (Abschnitt 3.6.1 Allgemeine Theorie der Schichtung) wurde darauf hingewiesen, daß zur Schichtung neben N_h-Werten der untersuchten Variablen auch Anteilssätze von Variablen genommen werden können, die mit der untersuchten Variablen eng korreliert sind. Hinsichtlich des Schweinebestandes stellen die Betriebe eine solche Variable dar. In der Stichprobe betrug die Korrelation zwischen dem Schweinebestand und der Anzahl der Betriebe in der Gemeinde

$$r(Y,M) = 0,91$$

Die Stichprobengemeinden werden also je nach Anzahl ihrer Betriebe in Schichten eingeteilt, so daß die Schichtanteile m_h der Stichprobe den n_h in Abschnitt 3.10.1.2.1 (Schichtung nach Größe des Schweinebestandes) etwa entsprachen. Die Gesamtzahl der Betriebe Belgiens wurde dann entsprechend dem Schichtungsmerkmal sortiert. Daraus erhielt man die M_h-Werte (Übersicht 4).

Übersicht 4 M_h-Werte der korrelierten Variablen

Schicht	Stichprobengemeinden, geschichtet nach Zahl der Betriebe m_h	Stichprobengemeinden, geschichtet nach Schweinebestand n_h	Zahl der Gemeinden in der Grundgesamtheit M_h
1	45	48	667
2	32	29	553
3	35	34	767
4	38	39	561
-	150	150	2548

Die zusätzliche Information über die Anteilssätze wurde nicht über die Schichtwerte des Schweine<u>bestandes</u>, sondern über die der <u>Betriebe</u> bestimmt. Die Parameter in Übersicht 1 ($\bar{y}_h$, s_h^2) werden nicht mit den N_h, sondern mit den M_h als zusätzliche Information hochgerechnet:

$$\hat{Y}_{p_3} = \sum_h M_h \bar{y}_h = 2\,742\,366 \text{ Schweine}$$

Die Formel für die Varianz (18) wird dann:

$$v(\hat{Y}_{p_3}) = \frac{N-n}{n}\left[\Sigma M_h s_h^2 + \frac{1}{n}\Sigma(M-M_h)s_h^2\right]$$

Daraus berechnet sich die Standardabweichung mit

$$\sqrt{v(\hat{Y}_{p_3})} = s_{\hat{Y}_{p_3}} = 296\,859 \text{ Schweine}$$

3.10.1.3 Vergleichende Darstellung

Wir vergleichen miteinander:

die Schätzwerte $\hat{Y}..$

die Standardabweichungen $s_{\hat{Y}..}$

die relativen Abweichungen $\Delta_{rel} = \frac{\hat{Y}..-Y}{Y}$

die relativen Standardabweichungen $s_{rel} = \frac{s_{\hat{Y}..}}{Y}$

den Grad der Wirksamkeit $\eta = \frac{s_{\hat{Y}}}{s_{\hat{Y}_{p..}}} 100$

Dabei ist Y der wahre Wert, den wir in unserem Beispiel zur Verfügung haben. Beim Wirksamkeitsgrad beziehen wir die Standardabweichung der einfachen Hochrechnung auf die Standardabweichungen der berechneten Fälle der A-posteriori-Schichtung. Der Wirksamkeitsgrad zeigt, wieviel wirksamer (in Prozent) die A-posteriori-Schichtung in den angeführten Fällen gegenüber der einfachen Hochrechnung ist.

Wichtig für die Beurteilung eines Schätzverfahrens ist weniger die Abweichung des Schätzwertes vom wahren Wert als vielmehr ein Vergleich der Varianzen. Die folgende Übersicht zeigt, daß die relativen Standardabweichungen (s_{rel}) bei den 3 Fällen der A-posteriori-Schichtung etwa 12% betragen, während sie bei der einfachen Hochrechnung fast das Doppelte, nämlich 21% ausmacht. Die letzte Spalte zeigt, daß die Wirksamkeit der A-posteriori-Schichtung etwa um 70% höher ist als bei der einfachen Hochrechnung.

Übersicht 5 Zusammenfassung der Ergebnisse

Schätzverfahren[1]	$\hat{Y}..$	$s_{\hat{Y}..}$	$\Delta_{rel}(\%)$	$s_{rel}(\%)$	$\eta(\%)$
einfache Hochrechnung	3 254 170	532 525	31,27	21,48	100
A-p-S mit "wahren" N_h	2 824 571	308 497	13,95	12,44	172,62
A-p-S mit geschätzten $\hat{N}_h$	2 973 346	317 951	19,94	12,83	167,49
A-p-S mit M_h einer korrelierten Variablen	2 742 366	296 859	10,62	11,97	179,39

[1]) A-p-S = A-posteriori-Schichtung

Es zeigt sich ferner, daß der Unterschied in der Wirksamkeit zwischen dem Verfahren mit den "wahren" N_h und jenem mit den geschätzten $\hat{N}_h$ nicht sehr groß ist. Das bedeutet, daß die Schichtenbildung mit Hilfe der Betaverteilung als A-priori-Verteilung ohne größeren Verlust der Wirksamkeit vorgenommen werden kann. Wenn man bedenkt, daß die Schätzwerte dabei nur aus den Stichprobendaten berechnet werden, erhält man mit diesem Verfahren gegenüber der einfachen Hochrechnung einen erstaunlichen Genauigkeitsgewinn.

Daß das Verfahren mit den M_h aus der korrelierten Variablen (Betriebe) etwas wirksamer ist als die Schichtung mit Hilfe der N_h (Schweinebestand), läßt sich vielleicht mit einer deutlicheren Schichtung bei den Betrieben als beim Schweinebestand erklären. Es mag sein, daß Industriegemeinden gegenüber bäuerlichen Gemeinden wenige aber große Betriebe aufweisen, so daß der Schweinebestand im ganzen eine kontinuierlichere Größe ist als die Betriebszahl.

3.10.2 Schätzung des Angabefehlers

In der Agrarstatistik können Angabefehler[1] bei offiziellen staatlichen Erhebungen, wie z.B. bei der Viehzählung, auf das Marktgeschehen einer Volkswirtschaft großen Einfluß haben. So kann eine dadurch verursachte Untererfassung des Viehbestandes eine erhöhte Einfuhr durch den Staat bewirken und damit die Verkaufspreise der Landwirte herunterdrücken. Aus diesem Grund wurde in Belgien versucht, die Größenklasse des Fehlers festzustellen, der durch die Falschangaben der Befragten entsteht.[2] Aus methodischen Gründen wurden hierfür zwei Auswahlmodelle mit variabler Auswahlwahrscheinlichkeit verwandt:

Modell B: Auswahl proportional zum Schweinebestand einer Gemeinde

Modell C: Auswahl proportional zur Anzahl der Betriebe einer Gemeinde

Auf die Fluktuation im Schweinebestand wurde schon ausführlich hingewiesen (Abschnitt 3.4.5). Aus diesem Grund ist hier eine A-posteriori-Schichtung geschickter als eine A-priori-Schichtung, da letztere an einem Material (z.B. Landwirtschafts- und Gartenbauzählung vom 15.Dez. 1964) vorgenommen werden muß, das für die Untersuchung (z.B. Landwirtschafts- und Gartenbauzählung vom 15.Mai 1965) doch einigermaßen veraltet ist.

Die Nachprüfung des Angabefehlers wurde stichprobenweise mit Hilfe der oben angegebenen Auswahlmodelle durchgeführt. Die Größe des Angabefehlers soll nun mit Hilfe der A-posteriori-Schichtung geschätzt werden. Dazu wollen wir uns auf ein einstufiges Modell beschränken, da dies für unsere Zwecke genügt.

Im Modell, das in Belgien angewandt wurde, wurden in jeder Gemeinde mit Hilfe der einfachen Zufallsauswahl 4 Betriebe ausgewählt. Dieses zweistufige Modell bauen wir in ein einstufiges um, indem wir den durchschnittlichen Fehler dieser 4 Betriebe als Untersuchungsmerkmal nehmen. Wir haben sozusagen einen durchschnittlichen Betriebsfehler der Gemeinde. Das tatsächlich angewandte Modell wird also für unsere Zwecke etwas umkonstruiert.

---

[1]) Angabefehler sind Fehler, die Betriebsinhaber bei der Deklaration ihres Viehbestandes begehen, sei es bewußt oder unbewußt.

[2]) S. Strecker, H./Steylaerts, R. (1966): S.55ff..

Die Fehler, die diese 4 Betriebe bzw. Betriebsinhaber begingen, erhielten wir durch die Nachprüfung der Schweinezählung in Belgien vom 15.Mai 1968. Die Fehler der 4 Betriebe in der Gemeinde werden nun summiert und durch 4 geteilt. Dann ist

$\bar{e}_i$ = wahrer (nachgeprüfter) Bestand - angegebener Bestand
= der Betriebsfehler in der i-ten Gemeinde

Diese Fehler, die wir durch die stichprobenweise Nachprüfung erhalten, werden bei der A-posteriori-Schichtung in Schichten aufgeteilt; somit

$\bar{e}_{hi}$ = Fehler der i-ten Gemeinde in der h-ten Schicht

Die Aufteilung in Schichten wurde so vorgenommen, daß bei beiden Modellen B und C aus Vergleichsgründen in jede Schicht ungefähr gleich viele Stichprobengemeinden fielen. Die Schichtaufteilung ergab sich dann beinahe zwangsläufig, da die Anzahl der Betriebe mit negativen Fehlern (angegebener Bestand größer als der tatsächliche) zu gering war, um eine eigene Schicht zu bilden, die Gruppe mit Null-Fehlern (keine Falschangaben) jedoch recht stark war. Diese beiden Gruppen wurden zusammengefaßt, so daß Schicht 1 stark besetzt war. Für die restlichen Schichten blieb dann eine annähernd gleich starke Besetzung (Übersicht 6).

Übersicht 6 Schichtaufteilung der Stichprobengemeinden nach Größe des Fehlers (Über- bzw. Unterangaben des Viehbestandes)

Schicht	Modell B		Modell C	
	Größe des festgestellten Fehlers $\bar{e}_{hi}$	Zahl der Gemeinden in der Schicht n_h	Größe des festgestellten Fehlers $\bar{e}_{hi}$	Zahl der Gemeinden in der Schicht n_h
1	-46 bis + 1	56	-21 bis 0	55
2	+ 2 bis + 7	33	+ 1 bis + 4	33
3	+ 8 bis +27	30	+ 5 bis +30	33
4	+28 und größer	31	+31 und größer	29
-	-	150	-	150

Die Anteilssätze in der Grundgesamtheit der nach Fehlern in Schichten aufgeteilten Gemeinden sind nicht bekannt. Es gab bisher keine Vollerhebung zur Feststellung des Angabefehlers, aus der die Schichtanteile in der Grundgesamtheit hergeleitet werden könnten. Deshalb greifen wir zur Schätzung der N_h auf die Beta-Funktion zurück, d.h., wir nehmen die Betaverteilung als A-priori-Verteilung (s. Abschnitt 3.6.2.2 und 3.10.1.2.2).

Mit Hilfe der Gleichungen (40), (37) und (35) (Abschnitt 3.6.2.2) berechnen wir zunächst die Hilfsparameter $\hat{\sigma}_p^2$, a und $\hat{p}_h$ (Übersicht 7).

Übersicht 7 Mit Hilfe der Betaverteilung berechnete Hilfsparameter

Parameter	Berechnete Werte für	
	Modell B	Modell C
$\hat{\sigma}_p^2$	0,003898	0,003429
a	11,7747	13,4202
$\hat{p}_1$	0,344	0,336
$\hat{p}_2$	0,227	0,228
$\hat{p}_3$	0,212	0,228
$\hat{p}_4$	0,217	0,208

Indem wir die $\hat{p}_h$ mit der Gesamtzahl der Gemeinden Belgiens multiplizieren, erhalten wir für die geschätzten $\hat{N}_h$:

Übersicht 8 Geschätzte $\hat{N}_h$-Werte

Schicht	Modell B		Modell C	
	n_h	$\hat{N}_h$	n_h	$\hat{N}_h$
1	56	876	55	856
2	33	579	33	581
3	30	540	33	581
4	31	553	29	530
-	150	2548	150	2548

Die Betaverteilung hat uns also die zusätzliche Information über die Schichtanteile der nach Fehlergröße geschichteten durchschnittlichen Betriebe pro Gemeinde geliefert. Die benötigten Formeln zur Hochrechnung für das Modell mit variabler Auswahlwahrscheinlichkeit liefert uns Abschnitt 3.7. Die transformierte Variable (43) ist

$$z_{hi} = \frac{\bar{e}_{hi}}{\hat{N}_h \pi_{hi}}$$

Dabei ist (42)

$$\pi_{hi} = \frac{p_{hi}}{p_h} \Sigma p_h$$

die Wahrscheinlichkeit, mit der die Werte hochgerechnet werden, und die bei der Berechnung der beiden Modelle jeweils definiert wird.

Der durchschnittliche Betriebsfehler in Belgien wird dann bei dem Modell der A-posteriori-Schichtung mit ungleicher Auswahlwahrscheinlichkeit (44):

$$\bar{e} = \sum_h^L \frac{N_h}{N} \frac{1}{n_h} \sum_i^{n_h} z_{hi}$$

Dabei ist der Schätzwert der Varianz (50)

$$v(\bar{e}) = \frac{N-n}{N \cdot n} \left[\sum_h^L \frac{N_h}{N} s_h^2 + \frac{1}{n} \sum_h^L \frac{N-N_h}{N} s_h^2 \right]$$

mit (49)

$$s_h^2 = \frac{\sum_i^{n_h} (z_{hi} - \bar{z}_h)^2}{n_h - 1} \quad .$$

Im folgenden wollen wir zunächst für beide Modelle getrennt die Berechnungen darstellen. Im letzten Abschnitt werden dann die Ergebnisse beider Modelle verglichen. Die Ausrechnungen wurden auf einer elektronischen Rechenanlage durchgeführt[1].

3.10.2.1 Modell B: Auswahl proportional zum Viehbestand

Zunächst müssen wir die Auswahlwahrscheinlichkeiten angeben. Y_i sei der Viehbestand in der Gemeinde und Y der Gesamtviehbestand in Belgien. Bei der Auswahl proportional zum Viehbestand (Modell B) ist die Auswahlwahrscheinlichkeit

$$p_i = \frac{Y_i}{Y}$$

und nach Aufteilung in Schichten

$$p_{hi} = \frac{Y_{hi}}{Y}$$

Die Wahrscheinlichkeit, mit der die Werte hochgerechnet werden (42), ist

$$\pi_{hi} = \frac{Y_{hi}/Y}{Y_h/Y} \frac{\Sigma Y_h}{Y} = \frac{Y_{hi}}{Y_h} \frac{\Sigma Y_h}{Y}$$

Die transformierte Variable (43) ergibt für das Modell B

$$z_{hi} = \frac{\bar{e}_{hi} Y_h Y}{N_h Y_{hi} \Sigma Y_h}$$

Beim Modell B führten die Berechnungen zu einem durchschnittlichen Angabefehler in Belgien von

$$\bar{e}_B = 2{,}658$$

mit einer Standardabweichung von

---

[1]) Hier danke ich Herrn K. Kafka, Institut für Wirtschafts- und Sozialstatistik der Universität Tübingen, für seine freundliche Mithilfe.

$$s_{\bar{e}_B} = \sqrt{v(\bar{e}_B)} = 0,436$$

3.10.2.2 Modell C: Auswahl proportional zur Anzahl der Betriebe

Wiederum müssen wir zunächst die Auswahlwahrscheinlichkeiten bei dem Modell angeben. M_i sei die Anzahl der Betriebe in einer Gemeinde und M die Gesamtzahl der Betriebe in Belgien. Bei der Auswahl proportional zur Anzahl der Betriebe (Modell C) ist die Auswahlwahrscheinlichkeit

$$p_i = \frac{M_i}{M}$$

und nach Aufteilung in Schichten

$$p_{hi} = \frac{M_{hi}}{M}$$

Die Wahrscheinlichkeit, mit der die Werte hochgerechnet werden, (42) ist

$$\pi_{hi} = \frac{M_{hi}/M}{M_h/M} \frac{\Sigma M_h}{M} = \frac{M_{hi}}{M_h} \frac{\Sigma M_h}{M}$$

Somit erhalten wir für die transformierte Variable (43) im Modell C

$$z_{hi} = \frac{\bar{e}_{hi} M_h M}{N_h M_{hi} \Sigma M_h}$$

Beim Modell C führten die Berechnungen zu einem durchschnittlichen Angabefehler in Belgien von

$$\bar{e}_C = 2,512$$

mit einer Standardabweichung von

$$s_{\bar{e}_C} = \sqrt{v(\bar{e}_C)} = 0,270$$

3.10.2.3 Vergleich beider Modelle

Wir dividieren den durchschnittlichen Fehler ($\bar{e}_B$, $\bar{e}_C$) und die Standardabweichung ($s_{\bar{e}_B}$, $s_{\bar{e}_C}$) durch den durchschnittlichen Schweinebestand je Betrieb:

$$\bar{Y}_M = \frac{Y}{M} = \frac{2\,479\,025}{90\,012} = 27{,}54$$

Zum Vergleich bilden wir folgende Größen:

$\bar{e}..{}_{rel} = \frac{\bar{e}..}{\bar{Y}_M} 100$ = Relativer durchschnittlicher Angabefehler (%)

$s_{\bar{e}..{}_{rel}} = \frac{s_{\bar{e}..}}{\bar{Y}_M} 100$ = Relative Standardabweichung des Angabefehlers (%)

$\hat{E}.. = Y \cdot \bar{e}..{}_{rel}$ = Schätzwert des Gesamtfehlers in Belgien (Zahl der zu wenig erfaßten Schweine)

$s_{\hat{E}..} = Y \cdot s_{\bar{e}..{}_{rel}}$ = Standardabweichung des Gesamtfehlers

Übersicht 9 Vergleich der Modelle B und C

Größe	Modell B		Modell C	
	mit A-posterio-ri-Schich-tung	ohne A-posterio-ri-Schich-tung	mit A-posterio-ri-Schich-tung	ohne A-posterio-ri-Schich-tung
$\bar{e}$	2,658	1,940	2,512	2,260
$s_{\bar{e}}$	0,436	0,476	0,270	0,372
$\bar{e}_{rel}$ %	9,652	7,044	9,120	8,207
$s_{\bar{e}_{rel}}$ %	1,584	1,729	0,979	1,352
E	239 284	174 626	226 099	203 459
s_E	39 274	42 873	24 279	33 516

Die Ergebnisse sind in Übersicht 9 zusammengefaßt. Um die Wirksamkeit der A-posteriori-Schichtung bei diesem Verfahren festzustellen, wurden die Modelle B und C vergleichsweise ohne A-posteriori-Schichtung berechnet; also einfache Hochrechnung bei Auswahl mit variabler Auswahlwahrscheinlichkeit.

Bei beiden Verfahren mit A-posteriori-Schichtung beträgt der durchschnittliche Angabefehler $\bar{e}$ etwas über 2,5, während er bei den Verfahren ohne A-posteriori-Schichtung stärker differiert. Dasselbe gilt auch für die relativen Fehler, die beim geschichteten Verfahren wesentlich weniger differieren als beim ungeschichteten.

Die vergrößerte Wirksamkeit des a-posteriori-geschichteten Verfahrens gegenüber dem einfachen Hochrechnungsverfahren zeigt neben der Varianz (Übersicht 9) der Variationskoeffizient. Der Variationskoeffizient ist definiert

$$V = \frac{s_{\hat{E}}}{\hat{E}} 100$$

Dieser ist bei den Verfahren

mit A-posteriori-Schichtung: $V_B = 19{,}4\%$ und
$V_C = 10{,}7\%$

ohne A-posteriori-Schichtung: $V_B = 25{,}6\%$ und
$V_C = 16{,}5\%$

Das zeigt, daß bei beiden Modellen mit Hilfe der A-posteriori-Schichtung die Wirksamkeit beträchtlich gestiegen ist.

4 Abschließende Bemerkungen

Zum Schluß sollen noch kurz die Kosten untersucht werden. Für die meisten Hochrechnungsverfahren und speziell für die A-posteriori-Schichtung ist in der Literatur über Kosten konkret nichts zu finden, obwohl das statistische Büro der Vereinten Nationen auf die Dringlichkeit von Kostenuntersuchungen und deren Veröffentlichung hinweist[1].

Bei der A-posteriori-Schichtung hat es wenig Sinn, eine allgemeine Kostenfunktion aufzustellen, etwa in der Art:

$$C = K + \sum_h c_h n_h$$

(K=konstante Kosten; c_h=Kosten der Schichteinheiten; n_h=Anzahl der Schichteinheiten in der Stichprobe)
Es geht auch nicht darum, das Minimum der Varianz bei gegebenen Kosten bzw. das Minimum der Kosten bei gegebener Varianz festzustellen[2]. Denn, da bei der A-posteriori-Schichtung die Auswahl schon vorliegt, können die Kosten vom Stichprobenumfang und von der Schichtanzahl als unabhängig betrachtet werden. Der geringe Mehraufwand an Programmier- und Rechenarbeit, wenn Schichtumfang und Schichtzahl bei der Ausrechnung variabel gehalten werden, fällt nicht ins Gewicht. Ein mehrmaliges Sortieren mit Hilfe einer elektronischen Rechenanlage ist kostenmäßig unerheblich. Im Rahmen einer Stichprobenerhebung werden die Hochrechnungskosten und somit auch die Kosten für die A-posteriori-Schichtung zu den fixen Kosten gerechnet[3]. Der Kostenfaktor "Hochrechnung" kann somit von den anderen Kostenfaktoren eliminiert werden.

Daher müßten statistische Institutionen ihre Kosten - aufgeteilt nach den verschiedenen Komponenten des Stichprobenplans (Vorbereitung, Auswahl, Hochrechnung, Auswertung) - veröffentlichen, damit Richtzahlen aufgestellt werden können, etwa wie bei einer Bauplanung, wo man bestimmte Faustregeln hat, die besagen, ein Projekt (Stichprobenuntersuchung) kostet in diesem Abschnitt

---

1) United Nations (Hrsg.)(1964): S.4/6.
2) Vgl. Hájek, Jaroslav (1958): S.130ff..
3) Vgl. Szameitat, Klaus/Schäffer, Karl-August (1964): S.125.

(Hochrechnung) in der Ausführung (A-posteriori-Schichtung) durchschnittlich diesen Betrag. Darüber hinaus ist es sehr schwierig, zur praktischen Seite der Kosten etwas zu sagen oder um mit Kish[1] zu sprechen: "Es ist sehr schwer, etwas Spezielles hierzu zu schreiben, ohne banal zu werden".

Was bei der A-posteriori-Schichtung von Vorteil ist, ist ja, daß Kostengesichtspunkte keine große Rolle spielen und daß der Statistiker bis zum Schluß die freie Konzeptwahl hat. Erst nach Vorliegen der Daten muß er sich für eine Schichtung entscheiden, und diese kann er bei minimalen Kosten ändern und neu berechnen. Man kann also laborieren, um bessere Ergebnisse zu erhalten.

In der vorliegenden Arbeit wurde versucht, die A-posteriori-Schichtung ziemlich erschöpfend zu behandeln, und die damit zusammenhängenden Probleme zu lösen. Nicht behandelt wurden mehrstufige Verfahren, da es sich hier von der A-posteriori-Schichtung aus gesehen um keine grundsätzlich neue Problematik handelt. Bei mehrstufigen Verfahren ist vor allem darauf zu achten, daß eine Schichtung in den oberen Stufen wegen der größeren Variabilität der Merkmale einen größeren Schichtungseffekt erbringt als eine Schichtung in den unteren Stufen[2]. Das heißt, bei einem zweistufigen Verfahren bringt eine nachträgliche Schichtung der ersten Stufe mehr an Effizienz als in der zweiten Stufe.

Es ist nicht ganz erklärlich, warum dieses Schätzverfahren in der Praxis so wenig Beachtung findet. Denn die größtmögliche Information erhalten wir, wenn wir jedes Merkmal einzeln betrachten, also sozusagen bei n Merkmalen n Schichten bilden. Da nun die einfache Stichprobe der "Einschichtfall" ist, somit der Fall, der die geringste Information liefert, ist es erstaunlich, daß die nachträgliche Schichten- bzw. Gruppenbildung der Stichprobendaten zur Erhöhung des Informationsgrades nicht öfters angewandt wird. Das Ziel der Statistik ist: genaue Information und präzise Kommunikation. Daher sollten alle An-

---

[1]) Kish, Leslie (1965): S.272.

[2]) Vgl. Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a): S.410.

strengungen unternommen werden, die Stichprobenpläne zu verfeinern (unter diesem Aspekt soll die vorliegende Arbeit einen Beitrag liefern), denn von den Ergebnissen der Statistiker hängen im täglichen Leben eine Reihe von Entscheidungen ab, und man kann ohne Übertreibung sagen: "Wir sind alle ein Opfer der Statistik"[1].

[1]) Freeman, Linton C. (1965): S.1

5 Literaturverzeichnis

Im Literaturverzeichnis werden die häufig auftretenden Zeitschriftentitel mit ihren Initialen angeführt:

JASA: Journal of the American Statistical Association;
AMS : Annals of Mathematical Statistics;
BISI: Bulletin of the International Statistical Institute;
JRSS: Journal of the Royal Statistical Society, wobei hier zwischen Serie A und B unterschieden wird.

American Statistician, The (1966/1967):

(Hrsg.: American Statistical Association)
1966 Vol.20 Nr.1, S.25; Nr.4, S.28; Nr.5, S.30;
1967 Vol.21 Nr.2, S.24; Nr.4, S.33.

Anderson, Oskar N. (1935):

Einführung in die mathematische Statistik, Wien, 1935.

Anderson, Oskar (1949):

Über die repräsentative Methode und deren Anwendung auf die Aufarbeitung der Ergebnisse der bulgarischen landwirtschaftlichen Betriebszählung vom 31.Dez. 1926, München, 1949.

Arkin, Herbert (1963):

Handbook of Sampling for Auditing and Accounting, Volume I - Methods, New York, San Francisko, Toronto, London, 1963.

Blumenthal, Saul (1968):

Multinomial Sampling with Partially Categorized Data, JASA, Vol.63, 1968, S.542-551.

Bowker, A.H./Lieberman, G.J. (1961):

Engineering Statistics, Englewood Cliffs, New York, Fourth Printing 1961.

Bowley, A.L. (1906):

Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science, JRSS, Bd.69, London, 1906, S.540-548.

Bryant, Edward C./Hartley, H.O./Jessen, R.J. (1960):

Design and Estimation in Two-way Stratification, JASA, Vol.55, 1960, Washington D.C., S.105-124.

Bulmer, M.G. (1965):

Principles of Statistics, Edinburgh, London, 1965.

Burnham, James (1949):

Die Machiavellisten, Verteidiger der Freiheit, Zürich, 1949.

Chernoff, Herman (1960):

A Compromise between Bias and Variance in the Use of Nonrepresentative Samples, Contributions to Probability and Statistics, Stanford University Press, Stanford, 1960.

Chiang, C.L. (1966):

On the Expectation of the Reciprocal of a Random Variable, The American Statistician, Vol.20, Nr.4, Washington, 1966, S.28.

Cochran, William G. (1959):

Sampling Techniques, New York, London, 1953, Second Printing, 1959.

Cochran, William G. (1961):

Comparison of Methods for Determining Stratum Boundaries, BISI, Bd.38,2, Tokyo, 1961, S.345-358.

Cochran, William G. (1964):

Sampling Techniques, Second Edition 1963, Second Printing 1964, New York, London.

Cox, D.R. (1957):

Note on Grouping, JASA, Vol.52, Washington D.C., 1957, S.543-547.

Cramér, Harald (1963):

Mathematical Methods of Statistics, Tenth Printing, Princeton, 1963.

Cureton, Edward E. (1968):

Unbiased Estimation of the Standard Deviation, The American Statistician, Vol.22, Nr.1, Washington, 1968, S.22.

Dalenius, Tore (1950):

The Problem of Optimum Stratification, Skandinavisk Aktuarietidskrift, Vol.33, Uppsala, 1950, S.203-213.

Dalenius, Tore (1957):

Sampling in Sweden, Contributions to the Methods and Theories of Sample Survey Practice, Stockholm, 1957.

Dalenius, Tore (1962):

Recent Advances in Sample Survey Theory and Methods, AMS, Vol.33, 1962, S.325-349.

Dalenius, T./Gurney, M. (1951):

The Problem of Optimum Stratification, II, Skandinavisk Aktuarietidskrift, Vol.34, Uppsala, 1951, S.133-148.

Dalenius, Tore/Hodges, Josef L. (1957):

The Choice of Stratification Points, Skandinavisk Aktuarietidskrift, Bd.40, Uppsala, 1957, S.198-203.

Dalenius, Tore/Hodges, Josef L. (1959):

Minimum Variance Stratification, JASA, Vol.54, Washington, 1959, S.88-101.

David, F.N./Barton, D.E. (1962):

Combinatorial Chance, London, 1962.

Deming, Edwards W. (1960):

Sample Design in Business Research, New York, London, 1960.

Deming, William Edwards (1966):

Some Theory of Sampling (New York, London, 1950, erste Ausgabe), New York, 1966 (Dover Edition).

Désabie, J. (1966):

Théorie et pratique des sondages, Paris, 1966.

Durbin, J. (1953):

Some Results in Sampling Theory when the Units Are Selected with Unequal Probabilities, JRSS, Series B, Vol.XV, London, 1953, S.262-269.

Durbin, J. (1958):

Sampling Theory for Estimates Based on Fewer Individuals than the Number Selected, BISI, Vol.36, III, Stockholm, 1958, S.113-119.

Durbin, J. (1959):

A Note on the Application of Quenouille's Method of Bias Reduction, Biometrika, Vol.46, London, Cambridge, 1959, S.477-480.

Durbin, J. (1959a):

Review of "Sampling in Sweden", JRSS, Vol.122, Series A, London, 1959, S.246-248.

Ekman, G. (1959):

An Approximation Useful in Univariate Stratification, AMS, Vol.30, 1959, S.219-229.

Fellegi, Ivan P. (1963):

Sampling with Varying Probabilities without Replacement: Rotating and Non-Rotating Samples, JASA, Vol.58, 1963, S.183-201.

Ferguson, Thomas S. (1967):

Mathematical Statistics, A Decision Theoretic Approach, New York, London, 1967.

Fisher, Ronald A. (1956):

Statistische Methoden für die Wissenschaft, 12.Aufl., Edinburgh, London, 1956.

Fisher, Walter D. (1953):

On a Pooling Problem from the Statistical Decision Viewpoint, Econometrica, Vol.21, Baltimore, Maryland, 1953, S.567-585.

Fisher, Walter D. (1958):

On Grouping for Maximum Homogeneity, JASA, Vol.53, Washington, D.C., 1958, S.789-798.

Freeman, Linton C. (1965):

Elementary Applied Statistics, New York, London, Sydney, 1965.

Freund, John E. (1962):

Mathematical Statistics, Englewood Cliffs, New York, 1962.

Fuller, Wayne A. (1966):

Estimation Employing Post Strata, JASA, Vol.61, 1966, S.1172-1183.

Geigy, J.R. A.G. (Hrsg.)(1960):

Documenta Geigy, Wissenschaftliche Tabellen, 6.Aufl., Basel, 1960.

Ghosh, Sakti P. (1963):

Estimating the Mean by Two-stage Sampling with Replacement, Calcutta Statistical Association Bulletin, Vol.12, 1963, S.97-103.

Ghosh, Sakti P. (1963a):

Postcluster Sampling, AMS, Vol.34, 1963.

Gini, Corrado (1958):

Logic in Statistics, Metron, Vol.19, Rom, 1958, S.1-77.

Goodman, Leo A. (1960):

On the Exact Variance of Products, JASA, Vol.55, Washington, 1960, S.708-713.

Goodman, Richard (1965):

Statistics, London, 1965.

Griffin, John I. (1962):

Statistics, Methods and Applications, New York, 1962.

Hagood, Margaret J./Bernert, Eleanor H. (1945):

Component Indexes as a Basis for Stratification in Sampling, JASA, Vol.40, Washington D.C., 1945, S.330-341.

Hájek, Jaroslav (1958):

Some Contributions to the Theory of Probability Sampling, BISI, Vol.36, Stockholm, 1958, S.127-134.

Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956a):

Sample Survey Methods and Theory, Vol.I, Methods and Applications, New York, London, 1956 (sec. printing).

Hansen, Morris H./Hurwitz, William N./Madow, William G. (1956b):

Sample Survey Methods and Theory, Vol.II, Theory, New York, London, 1956 (sec. printing).

Hartley, H.O./Rao, J.N.K. (1959):

Sampling with Unequal Probabilities and without Replacement, Proceedings of the Social Statistics Section, 1959 (American Statistical Association), S.56-59.

Hartley, H.O./Ross, A. (1954):

Unbiased Ratio Estimators, Nature, Vol.174, London, 1954, S.270-271.

Hauser, Siegfried (1965)(unveröff.):

Die verschiedenen Formen der Hochrechnung bei der Gewinnung statistischer Daten mit Hilfe der Hochrechnung, Tübingen, 1965, (unveröffentlichte Diplomarbeit).

Hendricks, Walter A. (1956):

The Mathematical Theory of Sampling, New Brunswick, New York, 1956.

Hendricks, Walter A. (1964):

Estimation of the Probability that an Observation Will Fall into a Specified Class, JASA, Vol.59, Washington D.C., 1964, S.225-232.

Hess, Irene/Sethi, V.K./Balakrishnan, T.R. (1966):

Stratification: A Practical Investigation, JASA, Vol.61, Washington D.C., 1966, S.74-90.

Horvitz, D.G./Thompson, D.J. (1952):

A Generalization of Sampling without Replacement from a Finite Universe, JASA, Vol.47, Washington D.C., 1952, S.663-685.

Hotelling, Harold (1960):

The Teaching of Statistics, Contributions to Probability and Statistics, Edited by Olkin, Ingram/Ghurye, Sudhish G./Hoeffding, Wassily/Madow, William G./Mann, Henry B., Stanford University Press, Stanford, 1960, S.11-24.

Jorgenson, D.W./McCall, J.J./Radner R. (1967):

Optimal Replacement Policy, Amsterdam, 1967.

Kapur, J.N./Saxena, H.C. (1963):

Mathematical Statistics, 3.Aufl., Delhi, New Delhi, Jullundur, Lucknow, Bombay, 1963.

Karlin, Samuel (1966):

A First Course in Stochastic Processes, New York, London, 1966.

Kellerer, Hans (1960):

Statistik im modernen Wirtschafts- und Sozialleben, Hamburg, 1960 (rde Nr.103/104).

Kellerer, Hans (1963):

Theorie und Technik des Stichprobenverfahrens, 3.Aufl., München, 1963.

Kiär, A.N. (1899):

Die repräsentative Untersuchungsmethode, Allgemeines Statistisches Archiv, Bd.5, Tübingen, 1899, S.1-22.

Kish, Leslie (1965):

Survey Sampling, New York, London, Sydney, 1965.

Krelle, Wilhelm (1962):

Verteilungstheorie, Wiesbaden, 1962.

Krishnaiah, Paruchuri R. (Hrsg.)(1966):

Multivariate Analysis, Proceedings of an International Symposium held in Dayton, Ohio, June 14-19, 1965, New York, London, 1966.

Kulldorff, Gunnar (1961):

Contributions to the Theory of Estimation from Grouped and Partially Grouped Samples, New York, Stockholm, 1961.

Lahiri, D.B. (1951):

A Method of Sample Selection Providing Unbiased Ratio Estimates, BISI, Vol.33/2, Kalkutta, 1951, S.133-140.

Lazar, Philippe (1966):

Partition d'un Groupe Hétérogène en Sous-Groupes Homogènes, Revue de Statistique Appliquée, Vol.XIV Nr.1, Paris, 1966, S.39-43.

Le Cam, L. (1964):

Sufficiency and Approximate Sufficiency, AMS, Vol.35, 1964, S.1419-1455.

Linder, Arthur (1960):

Statistische Methoden, 3. erw. Aufl., Basel, Stuttgart, 1960.

Lorenz, Paul (1951):

Die Sheppardsche Korrektur der Streuung, Mitteilungsblatt für Mathematische Statistik, Jahrgang 3, Heft 3, München, 1951, S.179-184.

Mack, C. (1966):

Essentials of Statistics for Scientists and Technologists, London, 1966.

Mahalanobis, P.C. (1946):

Recent Experiments in Statistical Sampling in the Indian Statistical Insitute, JRSS, Vol.109, London, 1946, S.325-370.

Mahalanobis, P.C. (1952):

Some Aspects of the Design of Sample Surveys, Sankhya, Vol.12, Calcutta, 1952, S.1-7.

Mauldon, I.G. (1959):

A Generalisation of the Beta-Distribution, AMS, Vol.30, 1959, S.509-520.

Mayet, P. (1905):

Stichprobenerhebungen in der Zwischenzeit zwischen großen Vollzählungen längerer Periodizität, BISI, Bd.XIV,2, Berlin, 1905, S.258-276.

McCrimmon, Kenneth R./Ryavec, Charles A. (1966):

An Analytical Study of the PERT Assumptions, Buffa, Elwood S., Readings in Production and Operations Management, New York, London, Sydney, 1966, S.519-540.

Menges, Günter (1959):

Stichproben aus endlichen Gesamtheiten, Theorie und Technik, Frankfurt, 1959.

Menges, Günter (1968):

Grundriß der Statistik, Teil 1: Theorie, Köln, Opladen, 1968.

Mertz, Friedrich Georg (1966):

Die Schätzung des Wertes eines Betriebslagers auf Stichprobenbasis, Allgemeines Statistisches Archiv, Bd.50/3, Göttingen, 1966, S.241-259.

Moran, P.A.P. (1968):

An Introduction to Probability Theory, Oxford, 1968.

Murthy, M.N. (1963):

Some Recent Advances in Sampling Theory,JASA, Vol.58, Washington D.C., 1963, S.737-755.

Nisselson, Harold (1950):

Anwendung von Stichprobenverfahren bei Zählungen, München, 1950.

Noelle, Elisabeth (1963):

Umfragen in der Massengesellschaft, Hamburg, 1963 (rde Nr.177/178).

O'Toole, A.L. (1964):

Elementary Practical Statistics, New York, London, 1964.

Pathak, P.K. (1964):

Sufficiency in Sampling Theory, AMS, Vol.35, 1964, S.795-808.

Peters, William S./Summers, George W. (1968):

Statistical Analysis for Business Decisions, Englewood Cliffs, 1968.

Råde, Lennart (1964):

Grouping and Combining, An Operations Research Problem, Skandinavisk Aktuarietidskrift, Jahrgang 1963, Uppsala, 1964.

Raj, Des (1963):

A Note on Stratification in Unequal Probability Sampling, Sankhya, Series B, Vol.25, Kalkutta, 1963, S.215/216.

Raj, Des (1966):

Some Remarks on a Simple Procedure of Sampling without Replacement, JASA, Vol.61, Washington D.C., 1966.

Randow, Thomas v. (1964):

Blauer Dunst, Rauchen und Krebs: Hier wird der Verzicht auf Kritik gefährlich, DIE ZEIT, Hamburg, 13.März 1964.

Rao, J.N.K. (1963):

On Three Procedures of Unequal Probability Sampling without Replacement, JASA, Vol.58, Washington D.C., 1963, S.202-215.

Rao, J.N.K./Hartley, H.O./Cochran, W.G. (1962):

On a Simple Procedure of Unequal Probability Sampling without Replacement, JRSS, Series B, Vol.24, 1962, S.482-491.

Rao, T.J. (1968):

On the Allocation of Sample Size in Stratified Sampling, Annals of the Institute of Statistical Mathematics, Vol.20, Nr.1, Tokio, 1968, S.159-166.

Richmond, Samuel B. (1964):

Statistical Analysis, Sec. Edition, New York, 1964.

Sampford, M.R. (1962):

An Introduction to Sampling Theory with Applications to Agriculture, Edinburgh, London, 1962.

Sampford, M.R. (1967):

On Sampling without Replacement with Unequal Probabilities of Selection, Biometrika, Vol.54, 1967, S.499-513.

Sasieni, Maurice/Yaspan, Arthur/Friedman, Lawrence (1959):

Operations Research: Methods and Problems, New York, London, Tokio, 1959.

Schmetterer, Leopold (1966):

Einführung in die mathematische Statistik, 2.erw. Aufl., Wien, New York, 1966.

Schott, S. (1917):

Das Stichprobenverfahren in der Städtestatistik (1917), Schott, S.: Ausgewählte Schriften, Mannheim, 1957, S.279-314.

Schott, S. (1923):

Statistik, 3.Aufl., Leipzig, Berlin, 1923.

Schreiter, Dieter/Stempell, Dieter (1968):

Kritischer Weg und PERT, 5.Aufl., Berlin, 1968.

Sethi, V.K. (1963):

A Note on Optimum Stratification of Populations for Estimating the Population Means, The Australian Journal of Statistics, Vol.5, 1963, S.20-33.

Slonim, Morris James (1960):

Sampling in a Nutshell, New York, 1960.

Smirnow, N.W./Dunin-Barkowski, I.W. (1963):

Mathematische Statistik in der Technik, Berlin, 1963.

Smirnow, W.I. (1959):

Lehrgang der höheren Mathematik, Teil III,2, 2.Aufl., Berlin, 1959.

Sokal, Robert R./Sneath, Peter H.A. (1963):

Principles of Numerical Taxonomy, San Francisko, London, 1963.

Solomon, Herbert (1966):

Jurimetrics, David, F.N. (Hrsg.): Research Papers in Statistics, Festschrift for J. Neyman, London, New York, Sydney, 1966.

Spiegel, Murray R. (1961):

Theory and Problems of Statistics, New York, 1961.

Springer, C.H./Herlihy, R.E./Mall, R.T./Beggs, R.I. (1966):

Statistical Inference, Homewood, 1966.

Statistisches Bundesamt (Hrsg.)(1954):

Statistische Berichte, Arbeits-Nr.V/35/1 vom 31.3.1954: Verkehrsleistung der Lastkraftwagen und Anhänger zur Lastenbeförderung in der Woche vom 7.-13.7.1952 und im Jahre 1952.

Statistisches Bundesamt (Hrsg.)(1960):

Stichproben in der amtlichen Statistik, Stuttgart, Mainz, 1960.

Statistisches Bundesamt (Hrsg.)(1966):

Das Arbeitsgebiet der Bundesstatistik, Stuttgart, Mainz, 1966.

Stephan, Frederick F. (1945):

The Expected Value and Variance of the Reciprocal and Other Negative Powers of a Positive Bernoullian Variate, AMS, Vol.XVI, Baltimore, 1945, S.50-61.

Stephan, Frederick F. (1948):

History of the Uses of Modern Sampling Procedures, JASA, Vol.43, Washington D.C., 1948, S.12-39.

Stephan, Frederick F./McCarthy, Philip J. (1958):

Sampling Opinions, An Analysis of Survey Procedure, New York, London, 1958.

Strecker, Heinrich (1957):

Moderne Methoden in der Agrarstatistik, Würzburg, 1957.

Strecker, H./Steylaerts, R. (1963):

Stichprobenverfahren in der belgischen Statistik der landwirtschaftlichen Arbeitskräfte, Statistische Informationen des Statistischen Amtes der Europäischen Gemeinschaften, Nr.4, 1963, S.5-70.

Strecker, H./Steylaerts, R. (1966):

Feststellung der Angabefehler bei Viehzählungen mit Hilfe von Stichproben, Statistische Informationen des Statistischen Amtes der Europäischen Gemeinschaften, Nr.2, 1966, S.55-102.

Stroschein, Fritz-Reinhard (1965):

Die Befragungstaktik in der Marktforschung, Wiesbaden, 1965.

Stuart, Alan (1958):

The Measurement of Estimation and Test Efficiency, BISI, Vol.36,3, 1958, S.79-86.

Stuart, Alan (1962):

Basic Ideas of Scientific Sampling, London, 1962.

Stuart, Alan (1964):

Some Remarks on Sampling with Unequal Probabilities, BISI, Vol 40,2, Toronto, 1964, S.773-780.

Sudman, Seymour (1966):

Probability Sampling with Quotas, JASA, Vol.61, Washington D.C., 1966, S.749-771.

Sukhatme, Pandurang V. (1963):

Sampling Theory of Surveys with Applications, New Delhi, Ames, 1963 (Erstdruck 1954).

Sukhatme, P.V. (1966):

Major Developments in Sampling Theory and Practice, David, F.N. (Hrsg.): Research Papers in Statistics, Festschrift for J. Neyman, London, New York, Sydney, 1966.

Szameitat, Klaus/Schäffer, Karl-August (1964):

Kosten und Wirtschaftlichkeit von Stichprobenstatistiken, Allgemeines Statistisches Archiv, Bd.48, Göttingen, 1964, S.123-146.

Szameitat, K./Zindler, H.J. (1966):

Automatic Detection and Correction of Errors, BISI, Vol.41,1, Belgrad, 1966, S.395-417.

Szulc, Stefan (1965):

Statistical Methods, Oxford, London, Edinburgh, New York, Paris, Frankfurt, 1965.

Tepping, Benjamin J./Hurwitz, William N./Deming, W. Edwards (1943):

On the Efficiency of Deep Stratification in Block Sampling, JASA, Vol.38, Washington D.C., 1943, S.93-100.

Thionet, M.P. (1953a):

La théorie des sondages, Paris, 1953.

Thionet, M.P. (1953b):

Applications des methodes de sondage aux enquêtes statistique, Paris, 1953.

Thionet, P. (1965):

Sur l'estimation de variance dans le cas d'échantillonage systematique, Revue de Statistique Appliquée, Vol.13, Nr.4, 1965, S.51-60.

Thionet, P. (1967):

Théorie des sondages: quelques problèmes recents, Journal de la Société de Statistique de Paris, 108.Jahrgang, 1.Trimestre, Paris, 1967, S.9-30.

Tschuprow, A.A. (1925):

Grundbegriffe und Grundprobleme der Korrelationstheorie, Leipzig, Berlin, 1925.

United Nations (Hrsg.)(1953):

Studies in Methods: Industrial Censuses and Related Enquiries, Vol.I: Text, Vol.II: Appendices, Series F, Nr.4, Statistical Office of the United Nations, New York, 1953.

United Nations (Hrsg.)(1955):

Studies in Methods: Handbook of Vital Statistics Methods, Series F, Nr.7, Statistical Office of the United Nations, New York, 1955.

United Nations (Hrsg.)(1960):

Studies in Methods: A Short Manual of Sampling, Series F, Nr.9, Statistical Office of the United Nations, New York, 1960.

United Nations (Hrsg.)(1964):

Recommendations for the Preparation of Sample Survey Reports (Provisional Issue), Series C, Nr.1, Rev.2, Statistical Office of the United Nations, New York, 1964.

Vance, Lawrence L./Neter, John (1956):

Statistical Sampling for Auditors and Accountants, New York, London, 1956.

Van der Waerden, B.L. (1957):

Mathematische Statistik, Berlin, Göttingen, Heidelberg, 1957.

Wadsworth, George P./Bryan, Joseph G. (1960):

Introduction to Probability and Random Variables, New York, Toronto, London, 1960.

Ward, Joe H. (1963):

Hierarchical Grouping to Optimize an Objective Function, JASA, Vol.58, Washington D.C., 1963, S.236-243.

Weibull, Christer (1960):

Some Aspects of Statistical Inference with Applications to Sample Survey Theory, Göteborg, 1960.

Weinberg, Franz (1968):

Grundlagen der Wahrscheinlichkeitsrechnung und Statistik sowie Anwendungen im Operations Research, Berlin, Heidelberg, New York, 1968.

Whitehead, A.N. (1958):

Eine Einführung in die Mathematik, 2.Aufl., Bern, München, 1958.

Williams, W.H. (1962):

The Variance of an Estimator with Post-Stratified-Weighting, JASA, Vol.57, Washington D.C., 1962, S.622-627.

Wine, Lowell R. (1964):

Statistics for Scientists and Engineers, Englewood Cliffs, New York, 1964.

Yamane, Taro (1967):

Elementary Sampling Theory, Englewood Cliffs, New York, 1967.

Yates, Frank (1965):

Sampling Methods for Census and Surveys, Third Edition (1960), Sec. Impression, London, 1965.

Yates, F./Grundy, P.M. (1953):

Selection without Replacement from within Strata with Probability Proportional to Size, JRSS, Series B, Vol.XV, 1953, S.253-261.

Yule, Udny G. (1927):

An Introduction to the Theory of Statistics, 8.Edition, London, 1927.

Zarkovich, S.S. (1960):

On the Efficiency of Sampling with Varying Probabilities and the Selection of Units with Replacement, Metrika, Bd.3, Würzburg, 1960, S.53-60.

Zarkovich, Slobodan S. (1961):

Sampling Methods and Censuses, Vol.I, Collecting Data and Tabulation, Rom, 1961.

Zarkovich, S.S. (1964):

Sampling Methods and Censuses, Vol.II, Quality of Statistical Data, Rom, 1964.

L e b e n s l a u f

Am 31.10.1940 wurde ich als Sohn des Fidel Hauser und seiner Ehefrau Viktoria geb. Hipp in Fischingen Kreis Hechingen geboren.

In Fischingen besuchte ich die Volksschule und in Horb a.N. von 1951 bis 1960 das Gymnasium, an dem ich am 11.März 1960 die Reifeprüfung ablegte. Anschließend studierte ich Volkswirtschaftslehre an den Universitäten Tübingen und Berlin und legte im Frühjahr 1965 an der Universität Tübingen die Diplomprüfung für Volkswirte ab.

Nach dem Diplomexamen war ich bis März 1968 als wissenschaftliche Hilfskraft und ab 1.3.1968 als Verwalter einer Assistentenstelle am Institut für Wirtschafts- und Sozialstatistik tätig.

Druck: Polyfoto Dr. Vogt KG
Stuttgart 1, Königstraße 51
Telefon 295681

Druck: Polyfoto Dr. Vogt KG
Stuttgart 1, Königstraße 51
Telefon 295681